



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Burns, Douglas

Title:

**Precision measurements of differential $t\bar{t}b\bar{b}$ production cross sections as a function
of kinematic event variables at 13 TeV at CMS**

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Precision measurements of differential $t\bar{t}$ production cross sections as a function of kinematic event variables at 13 TeV at CMS

By

DOUGLAS JOHN PAUL BURNS

Department of Physics
UNIVERSITY OF BRISTOL

IIHE
UNIVERSITY OF BRUSSELS



A dissertation submitted to the University of Bristol and Vrije Universiteit Brussel in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science, School of Physics, Bristol and Faculteit Wetenschappen en Bio-ingenieurswetenschappen, Vakgroep Fysica, Brussel.

December 13, 2018

Doctoral examination jury:

Prof. Dr. Jonas Rademacker
Prof. Dr. Jorgen D'Hondt
Dr. Miriam Watson
Dr. Rebeca Gonzalez Suarez

Supervisors:

Prof. Dr. Freya Blekman
Prof. Dr. Joel Goldstein
Prof. Dr. Greg Heath

Front cover illustration: A cross section of the open CMS detector.

M. Hoch and M. Brice, “CMS experiment”, (2008). CMS Collection

Rear cover illustration: A collision recorded on the first day of data taking in 2016.

T. Mc Cauley, “Collisions recorded by the CMS detector on 7 May 2016 at the start of the year’s physics run”, (2016). CMS Collection

ISBN number: 9789493079137

NUR Code: 924, 926

© 2018 Douglas John Paul Burns

All rights reserved. No parts of this book may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University of Bristol and Vrije Universiteit Brussels regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:

DATE:

Author's contribution

The research conducted for this thesis has been done in collaboration with the Top Quark Group at the University of Bristol and the Vrije Universiteit Brussel, and all material presented here, unless stated otherwise, has been conducted by the author and collaborators. The author has spent time studying at both institutions as well as spending nine months working at CERN.

Specific contributions of the author include a major update of the pre-existing analysis performed at Run-1, in order to expand and enhance it for data collected during Run-2. The author contributed to a preliminary result of differential top quark cross sections with respect to kinematic event variables, based on the first 71 pb^{-1} of 13 TeV data collected by CMS in 2015.

CMS Collaboration, “Measurement of differential top quark pair production cross sections in a fiducial volume as a function of event variables in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ”, CMS Physics Analysis Summary CMS-PAS-TOP-15-013, 2015 [7 cites]

The preliminary result incorporates several new kinematic event variables, however the small data sample led to measurements that were statistically dominated.

A new analysis, for which the author was the primary contact person and lead analyser, and which forms the basis of this thesis, was performed measuring the $t\bar{t}$ production cross sections with respect to the kinematic event variables used in the preliminary results with all the data collected by CMS in 2016. Many aspects were dealt with in more detail with respect to the preliminary results. The author was heavily involved in the update and implementation of systematic uncertainties, including the calculation of the b tagging efficiencies, the addition of a new set of systematic uncertainties based on the modelling in simulation and the creation of systematic covariance matrices used for the goodness-of-fit tests. The author added measurements of the absolute cross sections. The author has performed many validation tests both on data and simulation during the course of the research and has presented multiple times in CMS internal meetings, including the first of two approval talks for the analysis. The author wrote the bulk of the documentation of the analysis, including the paper. The paper is now published in the June 2018 volume of the Journal for High Energy Physics.

CMS Collaboration, “Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”, *JHEP* **06** (2018) 002, doi: 10.1007/JHEP06(2018)002, arXiv:1803.03991 [7 cites]

The author presented the findings of this research at the International Conference for High Energy Physics in July 2018, as well as summarising the state of top quark physics at CMS at that time. The author also presented this research at the 11th International Workshop on Top Quark Physics in September 2018. All work in this thesis, beyond the scope of the paper, has been performed solely by the author.

Besides the physics analysis, the author has contributed to the CMS experiment by performing Level-1 trigger shifts in the CMS control room during periods of data taking and was involved in efficiency studies for the High Level triggers used in 2016. Further experimental work is ongoing in relation to the so-called ‘Highly Ionising Particle’ effect. This is where a large deposition of charge can cause a loss in gain from the tracker readout chip, decreasing the tracker hit efficiency. The author created a standalone simulation to model the response of the readout chip with respect to the charge build up, with the hopes to supply a correction based on measured gain distributions. The problem, prevalent during data taking in the first half of 2016, was fixed by allowing the charge to bleed away quicker from the readout chip. The effect of charge build up may become an issue again in the future due to the much greater luminosities expected and so the work performed is still very relevant.

The author has performed outreach activities during the course of the research studies, in particular the CERN international Masterclass, the London Youth International Science Forum and giving lectures for high school students interested in Physics. The author has also been a demonstrator for degree students in the University of Bristol’s experimental laboratory.

Abstract

This thesis sets out to test the current understanding of the pair production of the top quark standard model. To do so, it first introduces the theoretical knowledge behind the standard model and the production of the top quark, before discussing how the top quark is modelled in simulation. Next, the large and complex machinery required to produce and detect high-energy collisions is described and the algorithm used to reconstruct particles from the deposits. These particles are used to perform measurements of the differential $t\bar{t}$ production cross sections. These are presented in the single-lepton (e or μ) decay channel, as a function of several kinematic event variables. These kinematic event variables do not require the reconstruction of the $t\bar{t}$ system and are N_{jets} , H_T , S_T , p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$. The measurements are performed with proton-proton collision data collected by the CMS experiment at the LHC during 2016, with an integrated luminosity of 35.9 fb^{-1} . The $t\bar{t}$ yield in each decay channel is calculated by subtracting the single top quark, vector boson in association with jets and multijet QCD production backgrounds from the total data yield. The $t\bar{t}$ yields are then extrapolated, using a $t\bar{t}$ model, with respect to the stable particles that form depositions detectable by CMS to a common phase space in order to be able to consistently combine the two channels. The data are also naturally smeared by the resolution of the detector and contain effects due to the detector efficiency and acceptance. The method to undo the thumbprint of the detector and to perform the extrapolation is discussed in detail. From the combined, unsmeared $t\bar{t}$ yields, both the normalised and absolute differential cross sections are calculated together with the statistical and systematic uncertainties. The cross section measurements are compared to state-of-the-art leading order and next-to-leading order $t\bar{t}$ simulations by means of a goodness-of-fit test, incorporating the total covariance matrix describing the correlations of the uncertainties between bins of the measurement. Finally, brief studies into the possible future of these differential cross sections measurements is discussed.

Contents

Figures and Plots

Tables

Acronyms

1	Introduction	1
2	Theoretical overview	5
2.1	The components of the standard model	5
2.2	Representing particles and their interactions	6
2.3	The strength of the QED interaction	8
2.4	Quantum chromodynamics	8
2.4.1	Running of the strong coupling and hadronisation	9
2.5	The weak interaction and electroweak unification	10
2.5.1	Electroweak symmetry breaking	12
2.5.2	The Cabbibo-Kobayashi-Maskawa mixing matrix	12
2.6	The shortcomings of the SM	13
2.6.1	Gravity	13
2.6.2	Massive neutrinos	13
2.6.3	Dark matter	14
2.6.4	Baryon asymmetry	14
2.6.5	Hierarchy problem	14
2.7	Top quark physics	14
2.7.1	Top quark production at the LHC	15
2.7.2	Top quark decay	19
2.7.3	Top quark cross section measurements	19
3	The Large Hadron Collider and the Compact Muon Solenoid	23
3.1	The Large Hadron Collider	23
3.1.1	A brief history of the LHC and CMS	23
3.1.2	Operating the LHC	26
3.2	Luminosity measurements	27
3.3	The Compact Muon Solenoid	29

3.3.1	Silicon pixel and strip trackers	30
3.3.2	Electromagnetic calorimeter	32
3.3.3	Hadronic calorimeter	34
3.3.4	Superconducting magnet	35
3.3.5	Muon detectors	37
3.3.6	Reducing the volume of data using triggering	38
3.3.7	Computational power	40
3.3.8	Data collection in 2016	40
4	Modelling proton-proton interactions	43
4.1	Hard interaction	43
4.2	Additional interactions and soft processes	45
4.3	Matching matrix-elements to parton showers	46
4.4	Fragmentation and colour reconnection	47
4.4.1	Lund string model	47
4.4.2	Cluster fragmentation model	49
4.5	Detector modelling	49
4.6	Production of top quark pair and background models	50
5	Reconstructing the physics objects	55
5.1	The particle flow algorithm	55
5.1.1	Particle flow elements	55
5.1.2	Linking algorithm	56
5.1.3	PF muons	57
5.1.4	PF electrons and photons	57
5.1.5	PF jets	58
5.1.6	PF missing transverse momentum	60
5.1.7	PF isolation	60
5.2	Analysis object selection	61
5.2.1	Electron objects	61
5.2.2	Muon objects	62
5.2.3	Jet Objects	62
6	Event selection and corrections	69
6.1	The event selection and kinematic event variables	69
6.2	Reweightings from the number of additional interactions	70
6.3	Reweightings from b jet identification	72
6.4	Reweightings from leptons	73
6.5	Multijet QCD	73
6.6	Data-simulation agreement in the signal region	88
6.6.1	Top quark \mathbf{p}_T	88
6.6.2	Calculating the yield of top quark pairs	91
7	Removing the CMS thumbprint	93
7.1	Particle level measurement	93
7.2	Choice of bins	94
7.3	Unfolding	101
7.4	Cross checking the unfolding	102
7.4.1	Checking the uncertainties	102
7.4.2	Checking for bias	105
7.4.3	Calculating the cross sections	111

8	Cross section uncertainties	113
8.1	Statistical uncertainties	113
8.2	Systematic uncertainties	113
8.2.1	Experimental uncertainties	114
8.2.2	Modelling uncertainties	115
8.3	Summarising the uncertainties	118
9	Differential cross section results	121
9.1	Cross section measurements	122
9.2	Goodness-of-fit tests	122
9.3	Effects of regularisation	124
10	Conclusions, prospects and outlook	135
10.1	Double differential cross sections	136
10.2	Additional event variables	143
10.3	Effective field theory	143
	Appendices	154
A	Purity and stability in the muon decay channel	155
B	Residual distributions	159
C	Response matrices in the muon decay channel	169
D	Regularisation strength in the muon decay channel	173
E	Pull distributions in the muon decay channel	177
F	Creating the envelopes to test for bias	181
G	Checking bias in the muon decay channel and in alternate models	187
H	Largest systematic uncertainties	195
I	Covariance Matrices	201
J	Differential cross sections as a function of lepton pseudorapidity	207
K	Additional double differential distributions	209
L	EFT studies	217
	Bibliography	223

Figures and Plots

1	A Feynman diagram showing the $t\bar{t}$ pair production by quark-antiquark annihilation and decay into the single lepton final state.	7
2	A set of Feynman diagrams probing the QED vertex at different energy scales. .	8
3	A Feynman diagram showing the strong interaction.	9
4	The running of the strong coupling constant.	10
5	A Feynman diagram showing the weak interaction.	11
6	Leading order Feynman diagrams for the pair production of top quarks.	15
7	Leading order Feynman diagrams for the production of single top quarks. . . .	16
8	The parton distribution function of the proton scaled by the fraction of momentum carried by the parton, $xf(x, \mu^2)$ with respect to the fraction of momentum carried by the parton.	17
9	A set of Feynman diagrams showing real NLO processes.	17
10	The inclusive $t\bar{t}$ cross section, measured by the ATLAS, CMS experiments and also at the Tevatron collider, in different decay channels at multiple \sqrt{s} . Also shown is the most precise theoretical prediction to date.	18
11	Comparisons of different $t\bar{t}$ models using established and new underlying event tunes.	21
12	The full CERN complex showing schematically the location of all the experiments on the ring complex, including the location of the four main experiments at the LHC: CMS, ATLAS, LHCb and ALICE.	24
13	A cross section of a LHC dipole magnet.	26
14	The integrated luminosities delivered by the LHC and recorded by the CMS experiment during 2016.	28
15	The integrated luminosity delivered by the LHC over all years of operation. . .	28
16	The distribution of the number of interactions in a bunch crossing.	29
17	A schematic of CMS, sliced to show the internal layout.	31
18	A slice through the CMS tracker detector in the $r - z$ plane.	32
19	A geometrical quarter view of the CMS ECAL.	33
20	The energy resolution of the CMS ECAL measured in a beam test.	34
21	The energy resolution of the CMS ECAL extracted from an unbinned likelihood fit on $Z \rightarrow e^+e^-$ events.	35
22	A geometrical quarter view of the CMS HCAL.	36

23	The transverse energy resolution of reconstructed jets in the CMS HCAL.	36
24	A geometrical one-quarter view of the CMS muon subsystems.	37
25	The muon p_T resolution as a function of p_T using the tracking only, muon sub-detectors only or both.	38
26	A schematic representing the process flow of the L1T.	39
27	An example simulation of a proton-proton collision showing all the various processes that must be modelled.	44
28	The four possible processes to strongly split a parton in the parton shower. . . .	45
29	A set of diagrams showing the possible ordering of the parton shower.	46
30	How the colour string is stretched by a gluon between the quark-antiquark colour dipole and how the fragmentation occurs for each string.	48
31	A sketch of the cluster model of fragmentation and hadronisation.	50
32	An example of clustering using the anti- k_T algorithm.	59
33	A schematic showing the factorised steps to correct the energy scale of a jet. . .	64
34	The simulated jet energy responses at 8 TeV and 13 TeV as a function of p_T and η	65
35	An illustration of a jet from a heavy flavour quark.	66
36	The flavour composition of jets from a $t\bar{t}$ model simulation.	67
37	Plots showing the distribution of the number of reconstructed primary vertices in simulation and data before and after pileup reweighting.	71
38	Plots showing the distribution of the number of reconstructed primary vertices in simulation and data, where σ_{minbias} has been varied by its uncertainty.	71
39	b tagging efficiencies with respect to jet p_T and η , depending on jet flavour. . . .	72
40	The b jet multiplicity distribution before and after reweighting.	73
41	The electron identification efficiency for data and simulation calculated with events with a Z boson present using the tag-and-probe method.	74
42	The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the non-isolated electron control region.	76
43	The distributions of the p_T^W , p_T^ℓ and $ \eta^\ell $ event variables and I_{rel} in the non-isolated electron control region.	77
44	The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the conversion electron control region.	78
45	The distributions of the p_T^W , p_T^ℓ and $ \eta^\ell $ event variables and I_{rel} in the conversion electron control region.	79
46	The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the non-isolated ($0.15 < I_{\text{rel}} < 0.30$) muon control region.	80
47	The distributions of the p_T^W , p_T^ℓ and $ \eta^\ell $ event variables and I_{rel} in the non-isolated ($0.15 < I_{\text{rel}} < 0.30$) muon control region.	81
48	The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables and I_{rel} in the non-isolated ($0.30 < I_{\text{rel}} < \infty$) muon control region.	82
49	The distributions of the p_T^W , p_T^ℓ and $ \eta^\ell $ event variables and I_{rel} in the non-isolated ($0.30 < I_{\text{rel}} < \infty$) muon control region.	83
50	The prediction of multijet QCD for the N_{jets} , H_T and S_T event variables in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region.	84

51	The prediction of multijet QCD for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region.	85
52	The prediction of multijet QCD for the N_{jets} , H_T and S_T event variables in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region.	86
53	The prediction of multijet QCD for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region.	87
54	The distributions of N_{jets} , H_T and S_T after full event selection.	89
55	The distributions of p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ after full event selection.	90
56	The distribution of the H_T variable after full event selection and the application of top quark p_T reweighting. With the application of the reweighting the ratio is now flat.	92
57	The purity and stability of the bins in the e+jets channel, measured using the POWHEG+PYTHIA simulation sample, for the N_{jets} , H_T and S_T event variables.	95
58	The purity and stability of the bins in the e+jets channel, measured using the POWHEG+PYTHIA simulation sample, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	96
59	The residual distribution, using events from both decay channels, for the H_T event variable.	97
60	The mapping between the reconstructed and particle-level distributions, using events from both decay channels, for the H_T event variable.	98
61	The set of migration matrices calculated in the e+jets channel, for the N_{jets} , H_T and S_T event variables.	98
62	The set of migration matrices calculated in the e+jets channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	99
63	A schematic view of the migration effects and statistical fluctuations.	101
64	A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the e+jets channel, for the N_{jets} , H_T and S_T event variables.	103
65	A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the e+jets channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	104
66	The pull distribution for the H_T event variable in the e+jets channel.	105
67	The pull mean and widths in relation to the bin numbers of the N_{jets} , H_T and S_T event variables in the e+jets channel.	106
68	The pull mean and widths in relation to the bin numbers of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the e+jets channel.	107
69	The H_T event distribution of the POWHEG+PYTHIA sample with the top quark p_T reweighted up and down to cover differences to data in the e+jets channel.	108
70	The cross sections and biases for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the e+jets channel.	109
71	The cross sections and biases for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the e+jets channel.	110

72	The colour reconnection uncertainties for the absolute differential cross section measurement as a function of the H_T event variable.	117
73	The normalised cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables with respect to different $t\bar{t}$ models.	124
74	The normalised cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ kinematic event variables with respect to different $t\bar{t}$ models.	125
75	The absolute cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables with respect to different $t\bar{t}$ models.	126
76	The absolute cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ kinematic event variables with respect to different $t\bar{t}$ models.	127
77	The normalised cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables with respect to different model parameters.	128
78	The normalised cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ kinematic event variables with respect to different model parameters.	129
79	The absolute cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables with respect to different model parameters.	130
80	The absolute cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ kinematic event variables with respect to different model parameters.	131
81	The $t\bar{t}$ yield after full event selection in the $e+\text{jets}$ and $\mu+\text{jets}$ channels as a function of the S_T and N_{jets} events variables.	137
82	The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the S_T and N_{jets} events variables, in the $e+\text{jets}$ (upper) and $\mu+\text{jets}$ (lower) channels.	138
83	Normalised and absolute double differential $t\bar{t}$ production cross section measurements with respect to the S_T and N_{jets} event variables in comparison to the POWHEG+PYTHIA model.	139
84	The $t\bar{t}$ yield after full event selection in the $e+\text{jets}$ and $\mu+\text{jets}$ channels as a function of the p_T^ℓ and $ \eta^\ell $ events variables.	140
85	The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the p_T^ℓ and $ \eta^\ell $ events variables, in the $e+\text{jets}$ (upper) and $\mu+\text{jets}$ (lower) channels.	141
86	Normalised and absolute double differential $t\bar{t}$ production cross section measurements with respect to the p_T^ℓ and $ \eta^\ell $ event variables in comparison to the POWHEG+PYTHIA model.	142
87	The distributions of $M_{b\ell}$, $\theta_{b\ell}$, $M3$, p_T^{j3j4} and additional p_T^{jet} after full event selection.	144
88	Feynman diagrams showing a vertex modified from effective field theory.	145
89	The upper panel shows the absolute differential cross sections with respect to the H_T event variable, predicted by setting the $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ parameter to -1 (red), 0 (green) and +1 (blue) TeV^{-2} respectively, compared to that measured in this thesis. The lower panel shows the $\Delta\chi^2$ value for EFT models of differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths and the absolute $t\bar{t}$ cross section with respect to the H_T event variable.	147
90	The purity and stability of the bins in the $\mu+\text{jets}$ channel, measured using the POWHEG+PYTHIA simulation sample, for the N_{jets} , H_T and S_T event variables.	156
91	The purity and stability of the bins in the $\mu+\text{jets}$ channel, measured using the POWHEG+PYTHIA simulation sample, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	157

92	The set of residual distributions for the first 7 bins of the H_T event variable calculated using the POWHEG+PYTHIA simulation.	160
93	The set of residual distributions for the last 6 bins of the H_T event variable calculated using the POWHEG+PYTHIA simulation.	161
94	The set of residual distributions for the first 7 bins of the S_T event variable calculated using the POWHEG+PYTHIA simulation.	162
95	The set of residual distributions for the last 6 bins of the S_T event variable calculated using the POWHEG+PYTHIA simulation.	163
96	The set of residual distributions for the p_T^{miss} event variable calculated using the POWHEG+PYTHIA simulation.	164
97	The set of residual distributions for the p_T^W event variable calculated using the POWHEG+PYTHIA simulation.	165
98	The set of residual distributions for the first 9 bins of the p_T^ℓ event variable calculated using the POWHEG+PYTHIA simulation.	166
99	The set of residual distributions for the last 8 bins of the p_T^ℓ event variable calculated using the POWHEG+PYTHIA simulation.	167
100	The set of residual distributions for the $ \eta^\ell $ event variable calculated using the POWHEG+PYTHIA simulation.	168
101	The set of migration matrices calculated in the μ +jets channel, for the N_{jets} , H_T and S_T event variables.	170
102	The set of migration matrices calculated in the μ +jets channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	171
103	A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the μ +jets channel, for the N_{jets} , H_T and S_T event variables.	174
104	A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the μ +jets channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	175
105	The pull mean and widths in relation to the bin numbers of the N_{jets} , H_T and S_T event variables in the μ +jets channel.	178
106	The pull mean and widths in relation to the bin numbers of the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the μ +jets channel.	179
107	The kinematic event distributions given by the POWHEG+PYTHIA sample reweighted up and down to cover any differences to data in the e+jets channel, for the N_{jets} , H_T and S_T event variables.	182
108	The kinematic event distributions given by the POWHEG+PYTHIA sample reweighted up and down to cover any differences to data in the e+jets channel for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	183
109	The kinematic event distributions given by the POWHEG+PYTHIA sample reweighted up and down to cover any differences to data in the μ +jets channel, for the N_{jets} , H_T and S_T event variables.	184
110	The kinematic event distributions given by the POWHEG+PYTHIA sample reweighted up and down to cover any differences to data in the μ +jets channel for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	185
111	The cross sections and biases for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the μ +jets channel.	188

112	The cross sections and biases for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the μ +jets channel.	189
113	The cross sections and biases for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the e+jets channel.	190
114	The cross sections and biases for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the e+jets channel.	191
115	The cross sections and biases for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the μ +jets channel.	192
116	The cross sections and biases for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables in the μ +jets channel.	193
117	The composition of the systematic uncertainties for the N_{jets} , H_T and S_T event variables.	196
118	The composition of the systematic uncertainties for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	197
119	The composition of the systematic uncertainties for the N_{jets} , H_T and S_T event variables.	198
120	The composition of the systematic uncertainties for the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	199
121	The total covariance matrix for the normalised differential cross section measurements with respect to the N_{jets} , H_T , and S_T event variables.	202
122	The total covariance matrix for the normalised differential cross section measurements with respect to the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	203
123	The total covariance matrix for the absolute differential cross section measurements with respect to the N_{jets} , H_T , and S_T event variables.	204
124	The total covariance matrix for the absolute differential cross section measurements with respect to the p_T^{miss} , p_T^W , p_T^ℓ and $ \eta^\ell $ event variables.	205
125	The normalised and absolute differential cross sections with respect to the $ \eta^\ell $ event variable, in the e+jets and μ +jets channels.	208
126	The $t\bar{t}$ yield after full event selection in the e+jets and μ +jets channels as a function of the S_T and p_T^{miss} events variables.	210
127	The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the S_T and p_T^{miss} events variables, in the e+jets (upper) and μ +jets (lower) channels.	211
128	Normalised and absolute double differential $t\bar{t}$ production cross section measurements with respect to the S_T and p_T^{miss} event variables in comparison to the POWHEG+PYTHIA model.	212

129	The $t\bar{t}$ yield after full event selection in the e+jets and μ +jets channels as a function of the p_T^{miss} and N_{jets} events variables.	213
130	The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the p_T^{miss} and N_{jets} events variables, in the e+jets (upper) and μ +jets (lower) channels.	214
131	Normalised and absolute double differential $t\bar{t}$ production cross section measurements with respect to the p_T^{miss} and N_{jets} event variables in comparison to the POWHEG+PYTHIA model.	215
132	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the N_{jets} event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	218
133	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the S_T event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	218
134	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the p_T^{miss} event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	219
135	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the p_T^{W} event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	219
136	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the p_T^ℓ event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	220
137	The left panel shows the predictions of EFT simulations using $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the $ \eta^\ell $ event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strengths.	220

Tables

1	The fermionic, half-integer spin, particles of the SM and their properties.	6
2	The bosonic, integer spin, particles of the SM and their properties.	6
3	The sets of single electron and single muon data collected during 2016 split into individual runs.	41
4	The set of simulated samples uused in this thesis.	52
5	The set of simulated samples associated with estimating the uncertainties in the POWHEG+PYTHIA modelling.	53
6	The identification criteria for signal and veto electrons.	62
7	The identification criteria for signal and veto muons.	63
8	The identification criteria for signal jets.	63
9	The transfer factors to the signal region from the four multijet QCD control regions.	75
10	The yield of $t\bar{t}$ estimated from the data in each bin for each event variable. . . .	91
11	A comparison of the bin width and the resolution in each bin for all event variables in the electon channel.	100
12	A comparison of the bin width and the resolution in each bin for all event variables in the muon channel.	100
13	The upper and lower bounds, in %, from each source of systematic uncertainty in the normalised differential cross section.	119
14	The upper and lower bounds, in %, from each source of systematic uncertainty in the absolute differential cross section.	120
15	Results of a goodness-of-fit test between the normalised cross sections in data and several models.	132
16	Results of a goodness-of-fit test between the absolute cross sections in data and several models.	132
17	Results of a goodness-of-fit test between the unregularised normalised cross sections in data and several models.	134

18	Results of a goodness-of-fit test between the unregularised absolute cross sections in data and several models.	134
19	Results of a goodness-of-fit test between the normalised cross sections as a function of $ \eta^\ell $ in data and several models, for the e+jets, μ +jets and e, μ +jets channels.	207
20	Results of a goodness-of-fit test between the absolute cross sections as a function of $ \eta^\ell $ in data and several models, for the e+jets, μ +jets and e, μ +jets channels.	207
21	The 95% confidence intervals reached for each event variable	217

Acronyms

ALICE	A Large Ion Collider Experiment.
ATLAS	A Toriodal LHC ApparatuS.
BEH	Brout-Englert-Higgs.
BSM	Beyond the Standard Model.
CDF	Collider Detector at Fermilab.
CERN	Conseil Européen pour la Recherche Nucléaire.
CKM	Cabbibo-Kobayashi-Maskawa.
CMS	Compact Muon Solenoid.
CP	Charge-Parity.
CR	Colour Reconnection.
CSC	Cathode Strip Chambers.
CSVv2	Combined Secondary Vertex (v2) algorithm.
D0	Tevatron experiment located in region D0.
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi.
DT	Drift Tubes.
ECAL	Electromagnetic Calorimeter.
ECFA	European Committee for Future Accelerators.
EFT	Effective Field Theory.
EM	Electromagnetic.
EWK	Electroweak.
FPGA	Field-programmable Gate Array.
FSR	Final State Radiation.
GSF	Gaussian Sum Filter.
HB	Hadronic Barrel detectors.
HCAL	Hadronic Calorimeter.
HERA	Hadron-Electron Ring Accelerator.
HE	Hadronic Endcap detectors.
HF	Hadronic Forward detectors.
HLT	High-Level Trigger.
HO	Hadronic Outer detectors.
ID	Identification.
IP	Impact Parameter.
ISR	Initial State Radiation.

JEC	Jet Energy Correction.
JER	Jet Energy Resolution.
JES	Jet Energy Scale.
L1T	Level-1 Trigger.
LEP	Large Electron-Positron collider.
LHCb	A Large Hadron Collider b experiment.
LHC	Large Hadron Collider.
LINAC2	Linear Accelerator 2.
LO	Leading Order.
MLM	M. L. Mangano.
MPI	Multiple Parton Interactions.
NLO	Next-to-Leading Order.
NNLO	Next-to-Next-to-Leading Order.
PDF	Parton Distribution Function.
PDG	Particle Data Group.
PF	Particle Flow algorithm.
PSB	Proton Synchrotron Booster.
PS	Proton Synchrotron.
PV	Primary Vertex.
QCD	Quantum Chromodynamics.
QED	Quantum Electrodynamics.
RIVET	Robust Independent Validation of Experiment and Theory.
RPC	Resistive Plate Chambers.
SC	Supercluster.
SM	Standard Model.
SNO	Sudbury Neutrino Observatory.
SPS	Super Proton Synchrotron.
SSC	Superconducting Super Collider.
SUSY	Supersymmetry.
SV	Secondary Vertex.
TEC	Tracker End Caps.
TIB	Tracker Inner Barrel.
TID	Tracker Inner Disks.
TOB	Tracker Outer Barrel.
WLCG	Worldwide LHC Computing Grid.
GEANT4	GEometry ANd Tracking 4.
HERWIG++	Herwig parton-shower model.
POWHEG	POsitive Weight Hardest Emission Generator matrix-element generator.
PYTHIA	Pythia parton-shower model.
MG5_aMC@NLO	MadGraph5_aMC@NLO matrix-element generator.

CHAPTER 1

Introduction

Since time immemorial, the people of our world have tried to answer the most fundamental questions we can ask. What is consciousness? Why do we dream? Are we alone in the Universe? How did life begin? How did the Universe begin? What is the Universe made of and how does it work? These are not easy questions. In truth, we can not definitively answer any of them. This does not, or ever will, dampen our spirits in the quest to answer them.

The standard model (SM) of particle physics [5–9] is our best attempt to describe our Universe on the smallest scale and how it evolves. It is an extremely successful, well tested model that describes the existence and behaviours of all the observed particles which form the constituents of the visible Universe. Indeed, it predicts accurately how particles interact at energy scales over many orders of magnitude; from simple magnetic compasses and powerful lasers to the high energy collisions experienced at current particle colliders.

It is, nevertheless, incomplete. It fails to include the simple action of an apple dropping in a gravitational field, it fails to describe more than the visible matter in the universe, a total of approximately 4% of its energy content, and it fails to describe why the matter that we do see is so much more prevalent than antimatter. These failures are seen as opportunities by physicists, however, searching for new particles and new interactions that can be used to describe Nature.

Perhaps one of the most exciting areas of the the SM to look at in detail is that related to the top quark. The top quark is the most massive, elementary (contains no substructure) particle of the SM and will decay without forming a hadron, and, as such, it allows us to study its bare quark properties through its decay products. Its large mass also means that the top quark is sensitive to many types of new physics, either by direct production, through its decay, or from interference produced by particles produced at higher energy scales. Aside from searching for new physics, the top quark appears in, and can be a major background to, many rare SM processes. Observing these processes will further test the SM. Many of the possible new physics processes will have a direct effect on the top quark pair ($t\bar{t}$) production cross section.

For these reasons and more, it is vital to know the $t\bar{t}$ production cross section to the highest degree of precision possible, not only for the inclusive (total) cross section, but as a function of variables allowing for studies into the shapes of the differential distributions. These differential measurements produce valuable information on how well the $t\bar{t}$ system is modelled and, if measured precisely enough, possible hints of new physics.

To obtain precision measurements, there needs to be a wealth of well-reconstructed $t\bar{t}$ data available. The Large Hadron Collider (LHC) is able to provide this data. It collides protons at an energy high enough for $t\bar{t}$ pairs to be produced easily and at an intensity high enough that they are produced frequently (≈ 30 M $t\bar{t}$ pairs produced during 2016). The Compact Muon Solenoid (CMS) experiment [10] provides exceptional resolution in both tracking and calorimetry, leading to excellent reconstruction of collisions and the production of a set of collision data eminently suitable for precise $t\bar{t}$ studies.

This thesis sets out to measure the differential cross sections of $t\bar{t}$ production with respect to several kinematic event variables. Event variables are special in that they do not require the full reconstruction of the full $t\bar{t}$ system, instead looking at the collision as a whole. This means that the measurements presented in this thesis as a function of the event variables are sensitive to not just the hard interaction producing the $t\bar{t}$ pair, but also the soft processes underneath, making them very suitable in the validation of event generation and modelling. Many of the event variables (missing momenta and jet activity) are also used as the variable of choice when searching for physics beyond the SM (BSM), for example in supersymmetry (SUSY). Precise measurements of the cross sections for one of the largest backgrounds in these searches, with respect to their search variables will be very useful. The event variables can also be designed as a proxy for the top quark, without having to reconstruct it, providing additional information on it.

The measurements are performed using collision data, collected by the CMS experiment in 2016, with a single lepton (either electron or muon) in the final state. This is because of the large branching ratio for $t\bar{t}$ decaying into this final state and relatively small contributions of the background processes mimicking this final state. In order to combine the electron and muon channel consistently, the measurements are presented in a common phase space, similar to that accessible by the CMS experiment, and with respect to stable particles that can be detected. This is done through the process of unfolding, which also takes into account the efficiency, acceptance and resolution effects of the detector at the same time. The unfolding removes the thumbprint of the CMS detector from the measurements, meaning they can be easily compared to other $t\bar{t}$ production models without having to go through the time-expensive step of detector modelling and also to results from other experiments.

Both absolute and normalised differential cross sections are calculated as a function of the event variables. By measuring the absolute differential cross sections a maximal amount of information is provided on the cross sections and their associated uncertainties. By normalising the differential cross sections with respect to the total cross section in the visible phase space, some uncertainties which are highly correlated between bins of the measurement can be reduced significantly. This allows for precise studies on the shape of the distributions.

It is widely expected for these measurements to be used in the tuning of future $t\bar{t}$ generators. In addition, these measurements can be used for constraints on the PDF of the proton, as an accurate background estimate for BSM or rare SM searches and as an input to phenomenology studies.

It is worth noting that natural units are used in this thesis where $c = \hbar = 1$, meaning that mass, momentum and energy are all measured using eV. A summary of the content presented in this thesis is outlined here:

- Ch. 2 - A brief summary of the SM, its shortcomings and top quark physics.
- Ch. 3 - A discussion on the LHC, the CMS experiment and the data collection performed.
- Ch. 4 - How interactions are modelled in simulation and a description of the simulated datasets used in this thesis.

- Ch. 5 - The events have been collected in data and modelled in simulation and now they must be reconstructed into usable analysis objects.
- Ch. 6 - The event selection is defined and the reconstructed $t\bar{t}$ yields calculated.
- Ch. 7 - The detector effects are removed, decay channels combined and differential cross sections defined.
- Ch. 8 - The statistical and theoretical uncertainties on the differential cross sections are investigated.
- Ch. 9 - The normalised and differential cross sections are presented.
- Ch. 10 - The possibilities to extend the differential cross section measurements.

CHAPTER 2

Theoretical overview

The SM is an elegant attempt to describe the fundamental nature of the Universe. At the heart of the model are the base constituents of matter, the elementary particles, and the interactions between them, the forces. The SM describes the current experimental measurements well across a large range of energy scales. It manages to describe forces that we can see in the macroscopic world to those binding together the constituents of protons. The SM is an incomplete model, however. It does not, for example, describe the gravitational interactions between particles, nor does it fully explain the matter-antimatter imbalance seen in the Universe. Previous discrepancies in the SM have lead to predictions and discoveries in Nature, such as the existence of three generations of quarks due to CP-violation, first observed in neutral kaon decays [11]. In this way, the SM has evolved over the half-century since it was first postulated, to incorporate new experimental evidence to become a more complete description of nature and it will continue to do so for many years to come. This chapter will discuss the particles which make up the SM and how they interact with each other. Some drawbacks of the SM will then be highlighted, before moving on to the physics of top quarks and their cross sections.

2.1 The components of the standard model

The SM contains twelve types of fermions, split into leptons (ℓ) and quarks (q), whose interactions are mediated by four types of gauge boson. The set of leptons exists in three generations: electrons, muons and taus, together with their associated neutrinos: (ν_e, e) , (ν_μ, μ) and (ν_τ, τ) . The set of quarks (q) also exist in three generations: up and down (u, d), charm and strange (c, s) and top and bottom (t, b). For each fermion (f) that exists there is an antifermion (\bar{f}), with identical quantum properties except conjugated charges. All fermions are spin $1/2$ particles. The properties of the fermionic particles are shown in Tab. 1, as listed by the Particle Data Group (PDG) in 2016 [12].

Particles in the SM can interact via the electromagnetic (EM), weak and strong forces. In particle physics, these forces are described by quantum field theories, in which the interactions between fermions are mediated by the exchange of a virtual boson. The EM force is mediated by the photon (γ), the weak force by the massive W^\pm and Z bosons and the strong force by a set of eight gluons (g). These gauge bosons are all integer spin particles.

Table 1: The fermionic, half-integer spin, particles of the SM and their properties. They are split into three generations of leptons and quarks.

Generation	Leptons			Particle	Quarks	
	Particle	Mass (MeV)	Charge (e)		Mass (MeV)	Charge (e)
1	ν_e	$< 2 \times 10^{-6}$	0	u	2.2	$2/3$
	e	0.511	-1	d	4.7	$-1/3$
2	ν_μ	< 0.19	0	c	1270	$2/3$
	μ	105.7	-1	s	96	$-1/3$
3	ν_τ	< 18.2	0	t	173200	$2/3$
	τ	1777	-1	b	4180	$-1/3$

Table 2: The bosonic, integer spin, particles of the SM and their properties. The force mediators are spin 1 and the H boson spin 0.

Force Mediator	Particle	Mass (GeV)	Charge (e)
EM	γ	0	0
Weak	W	80.385	± 1
	Z	91.188	0
Strong	g	0	0
—	H	125.09	0

The last particle of the current SM is the spin 0 Brout-Englert-Higgs (BEH) boson (H), which is an excitation of the BEH field. The interaction with the BEH field, via the BEH mechanism, is responsible for imparting mass to the fundamental particles [13–15]. The properties of the bosons are shown in Tab. 2 [12].

While the SM is able to describe three of the four fundamental forces between particles, it is not a complete model. It is unable to reconcile gravity with the other forces at a fundamental particle level. This is one of the discrepancies seen between the SM and nature, and must be one of the future evolutions of the SM. Other problems include the baryon asymmetry seen in the Universe, the requirement for dark matter and the hierarchy problem. These are discussed in Sec. 2.6.

2.2 Representing particles and their interactions

Describing the dynamics of particles and their interactions with other particles through the excitations of relativistic quantum fields is not trivial. In 1948, Richard Feynman developed a powerful, diagrammatic tool [16, 17], to help with the computation of these interactions. This diagrammatic tool is now commonly called the Feynman diagram. All Feynman diagrams in this thesis are created using the feyntikz package [18]. A complete Feynman diagram showing a complex example of top quark-antiquark ($t\bar{t}$) production and decay is shown in Fig. 1. For all Feynman diagrams shown in this thesis, time increases from the left-hand side to the right, such that all the initial state particles are represented on the left and final state particles towards the right. Each arrowed line in the Feynman diagram represents a fermionic particle of the SM, each sinusoidal line represents an electroweak mediator particle (W^\pm , Z, γ) and each coiled line a gluon. Anti-fermions are represented as fermions moving backwards in time. Free fermions and antifermions are described by the Dirac equation, which can be obtained by substituting the Lagrangian density for a Dirac field into the Euler-Lagrange equation [19]. The Lagrangian density for free fermions is given by:

$$\mathcal{L}_{\text{free}}(x) = \bar{\Psi}(x)(i\gamma^\mu\partial_\mu - m)\Psi(x), \quad (2.1)$$

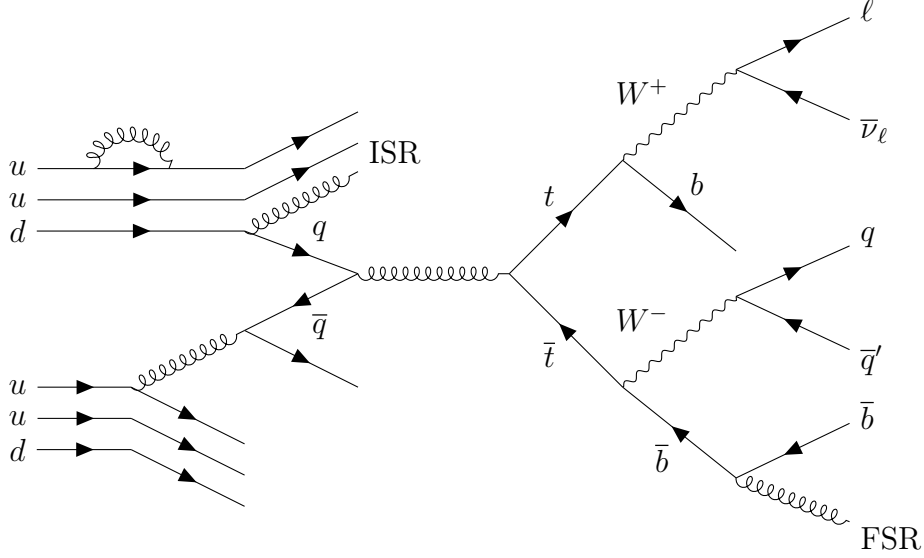


Figure 1: A Feynman diagram showing the $t\bar{t}$ pair production at the LHC by quark-antiquark annihilation and decay into the single lepton final state. Also shown are radiative processes in both the initial (ISR) and final state (FSR). Within the proton, there are also many interactions occurring, such that the colliding partons in the initial state are described by a parton distribution function.

where $\Psi(x)$ is the fermionic field depending on space-time, γ^μ are the gamma matrices, ∂_μ is the partial derivative and m the fermion mass. The laws of physics are invariant under transformation, e.g. they are the same at every point in space. Applying a global phase transformation,

$$\Psi(x) \rightarrow \Psi'(x) = e^{iq\chi}\Psi(x), \quad (2.2)$$

which is a global U(1) symmetry, to $\mathcal{L}_{\text{free}}$ leaves $\mathcal{L}_{\text{free}}$ unchanged. By Noether's theorem [20] there must be a conserved current and charge, associated to a global symmetry, which can in this case be identified as the EM current and electric charge. When a local phase transformation is applied however,

$$\Psi(x) \rightarrow \Psi'(x) = e^{iq\chi(x)}\Psi(x), \quad (2.3)$$

where the phase $q\chi(x)$ depends on space-time, the Lagrangian is no longer invariant, but contains the additional term:

$$\mathcal{L}(x) = \mathcal{L}_{\text{free}}(x) - q\bar{\Psi}(x)\gamma^\mu\partial_\mu(\chi(x))\Psi(x). \quad (2.4)$$

In order to maintain local invariance, $\mathcal{L}_{\text{free}}$ needs to be modified to produce a contribution that exactly cancels the additional term gained from the transformation. A new gauge vector field, A_μ , and coupling constant, g , is introduced into the Lagrangian density, via the covariant derivative

$$D_\mu = \partial_\mu + igA_\mu. \quad (2.5)$$

This gauge field, the photon field, interacts with the fermion field, Ψ , cancelling the additional term and maintaining gauge invariance, provided the photon field is massless. The combination of the free fermionic fields, photon fields and the EM interaction term is known as Quantum Electrodynamics (QED), with the full Lagrangian density given as:

$$\mathcal{L}_{\text{QED}} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + g\bar{\Psi}\gamma^\mu\Psi A_\mu, \quad (2.6)$$

where $F^{\mu\nu}$ is the EM field strength tensor with the term describing free photons,

$$F_{\mu\nu} = \partial_\mu F_\nu - \partial_\nu F_\mu. \quad (2.7)$$

2.3 The strength of the QED interaction

The strength of the QED interaction can be experimentally measured, however the value obtained is not the same as the bare coupling constant seen in Feynman diagrams. It is the effective strength of the interaction, including all contributions from relevant higher-order QED corrections, for example including those seen in Fig. 2. The higher-order corrections introduce loops into the Feynman diagram, with each loop introducing an integral over the four-momenta of all particles present in that loop. The loops lead to divergent corrections and infinities being present.

The diagrams in the lower panels of Fig. 2 represent corrections to the electron four-vector current and exactly cancel to all orders of perturbation theory. This leaves only the corrections to the photon propagator which can be folded into the definition of the coupling constant, giving the coupling constant, $g(\mu^2)$, a dependence on the energy scale it is measured at, μ . This dependence on energy is known as the running of the coupling. The strength of the EM coupling constant increases with the energy scale due the screening of the electric charge by the additional lepton-antilepton pairs. The energy scale at which $g(\mu^2)$ is evaluated at is known as the renormalisation scale.

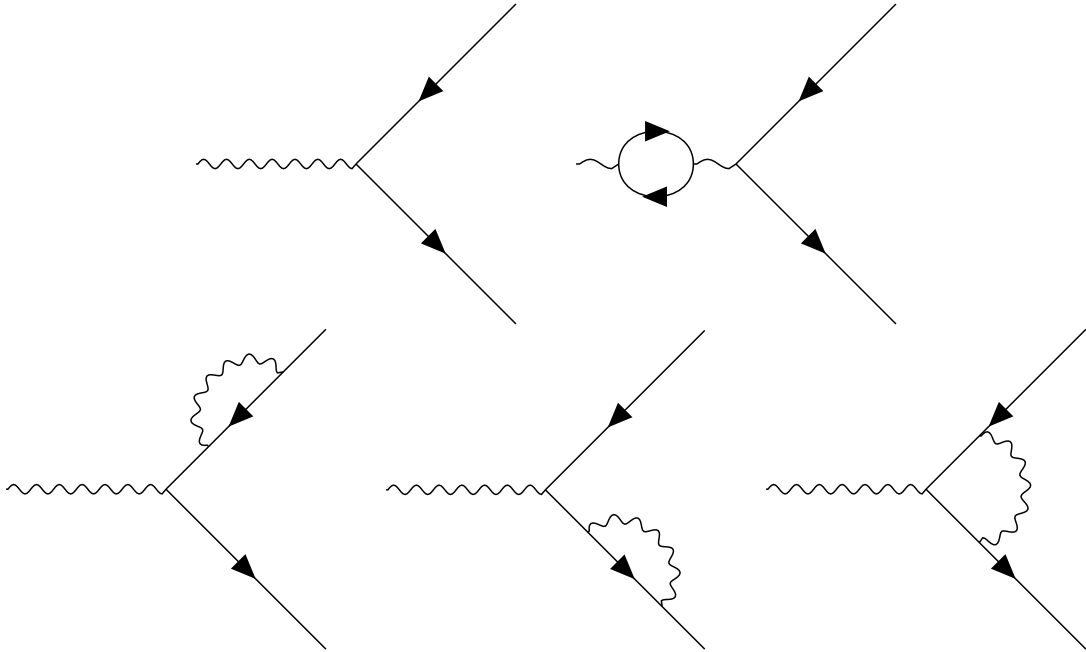


Figure 2: A Feynman diagram showing a QED vertex. The left panel shows a simple lepton-antilepton pair production, however as the energy scale increases virtual particles contribute to the interaction vertex, leading to the running of the coupling constant.

2.4 Quantum chromodynamics

The quantum field theory for the strong interaction, Quantum Chromodynamics (QCD), is similar in many ways to that of QED. The strong interactions in the $t\bar{t}$ system are highlighted in red on Fig. 3. It is described by an $SU(3)$ symmetry group, with three orthogonal states known as colour charges (r, g, b). To maintain invariance under the local $SU(3)$ gauge transformation

$$\Psi(x) \rightarrow e^{i\theta^a(x)t^a}\Psi(x), \quad (2.8)$$

eight new gauge fields, G_μ^a , must be introduced in a similar manner to QED. These fields come from the eight generators of the $SU(3)$ symmetry group t^a , which can be represented by the

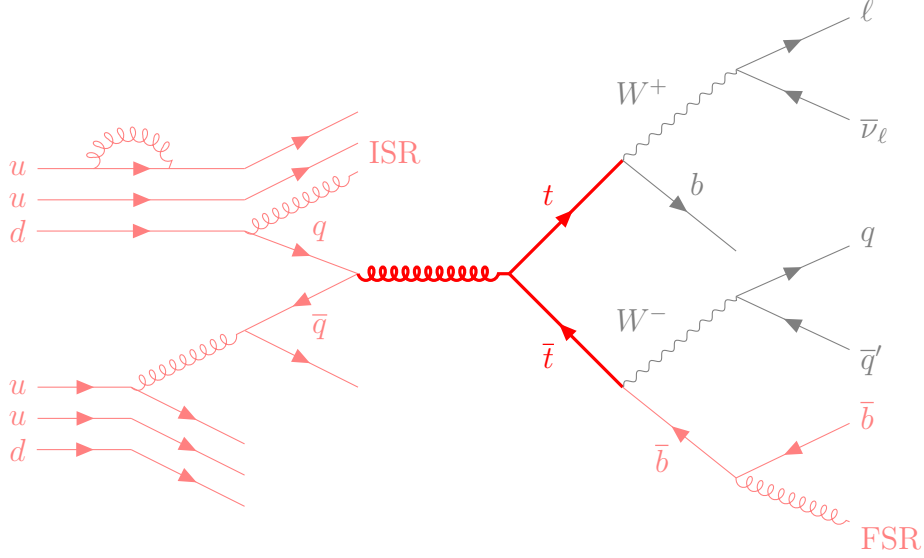


Figure 3: A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example QCD interaction is shown in bold red. Other QCD interactions are shown in light red.

Gell-Mann matrices as $t^a = -\frac{1}{2}\lambda^a$. The λ^a matrices are traceless and Hermitian forming the basis of the $SU(3)$ Lie algebra.

As with QED, the extra fields are folded into the covariant derivative, this time with the strong coupling constant, g_s ,

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a. \quad (2.9)$$

Thus, the strong force is mediated by eight massless gauge bosons (gluons), however, in contrast to QED, these gluons carry a colour charge brought about by the non-commutation of the $SU(3)$ group generators, leading to gluon-gluon self-interactions. The full QCD Lagrangian density is given as

$$\mathcal{L}_{\text{QCD}} = \bar{\Psi}_i ([i\gamma^\mu D_\mu]_{ij} - m\delta_{ij}) \Psi_j - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \quad (2.10)$$

where the gluon field strength tensor

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \quad (2.11)$$

is analogous to the EM field strength tensor with the $SU(3)$ fine structure constants, f^{abc} , introduced by the gluon-gluon self-interaction.

2.4.1 Running of the strong coupling and hadronisation

Renormalisation in QCD is treated similarly as in QED, except that there are additional loop diagrams introduced because of the self-interaction of gluons. The corrections due to gluon loops are of the opposite sign as those introduced from the fermion loops. This means that the strong coupling constant runs in the opposite direction as the QED coupling constant, i.e. at higher energies g_s (or $\alpha_S = g_s^2/4\pi$) tends to 0. This is known as asymptotic freedom and is shown in Fig. 4.

As can be seen, α_S varies considerably over the large range of energies present in particle physics. At higher energy scales, for example at $\mu = 100$ GeV, α_S is sufficiently small for perturbation theory to be valid. This means that in collider experiments, the quarks inside the proton can be considered to be quasi-free particles, instead of being strongly bound. It is not sufficiently small enough, however, that higher order corrections can be ignored. This means

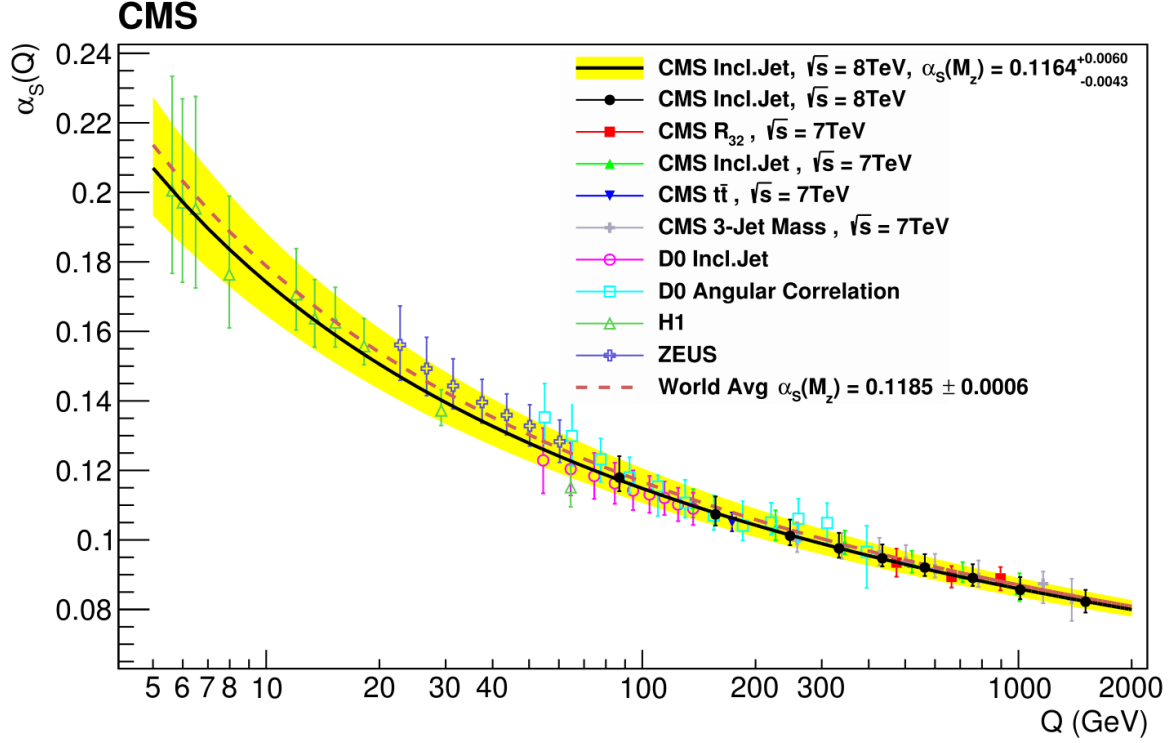


Figure 4: The running of the strong coupling constant. At higher energy scales α_s tends to 0. Figure taken from [21].

any QCD calculation is always carried out beyond leading order, which is extremely difficult and involves many Feynman diagrams.

At an energy scale $\mu \approx 1$ GeV, α_s is of order unity and perturbation theory breaks down. This energy scale is often referred to as Λ_{QCD} . The colour confinement hypothesis states that particles with colour charge cannot propagate as a free particle, and so, at below Λ_{QCD} , they exist in bound, composite, colourless states, known as baryons (qqq) and mesons (q \bar{q}).

The processes of fragmentation and hadronisation describe how the final state quarks and gluons, as seen in Fig. 1, end up as baryons and mesons. Fragmentation is where the particles are initially travelling away from each other at high velocities, and as they separate the colour field between them stretches. As the colour field stretches, the energy stored within it increases until it is favourable to produce a new q \bar{q} pair, breaking the colour field into two. This continues until all the quarks have sufficiently low energy, $< \Lambda_{\text{QCD}}$, to form colourless hadrons (hadronisation). The hadronisation process occurs within a time scale of $\tau_{\text{had}} \simeq 1/\Lambda_{\text{QCD}} \approx 3 \times 10^{-24}$ s and forms a spray of hadrons in the direction of the initiator quark or gluon and is known in experimental terms as a jet. If an interaction occurs from particles within a hadron, for example in proton-proton collisions, then the hadron ceases to be a colourless state, and the remnants hadronise forming more sprays of particles. These particles are known as the underlying event. The modelling of the fragmentation and hadronisation processes will be discussed in more detail in Sec. 4.4.

2.5 The weak interaction and electroweak unification

The weak interaction differs from QED and QCD in that it is possible to violate parity conservation and to change the flavour of a quark. Example weak interactions in the $t\bar{t}$ system are highlighted in blue on Fig. 5. The weak interaction is described by an SU(2) symmetry group,

with the conserved charge being the weak isospin, T . The third component of weak isospin T_3 , is also conserved. Parity violation is seen in the weak interaction as only left-handed particles and right-handed antiparticles have been observed to interact via the weak force. At higher

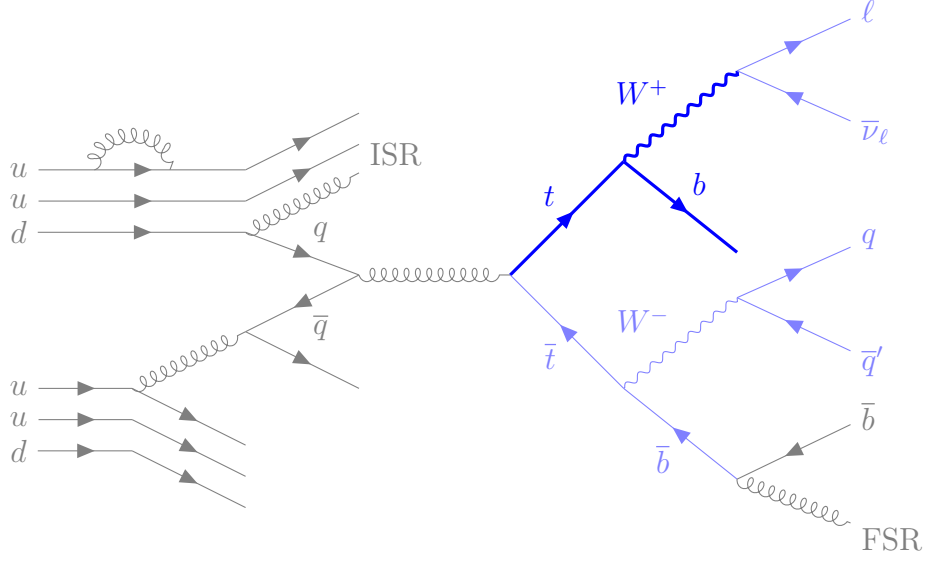


Figure 5: A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example weak interaction is shown in bold blue. Other weak interactions are shown in light blue.

energies, the electromagnetic and weak interactions unify into the electroweak (EWK) interaction. This combination results in a $SU(2)_L \otimes U(1)_Y$ symmetry group, with three gauge fields, W_μ^i , introduced by the $SU(2)_L$ generators $\frac{1}{2}\sigma_i$ (the Pauli spin matrices), and one gauge field, B_μ , from the $U(1)_Y$ generator (weak hypercharge $Y_W = 2(q - T_3)$). The EWK gauge bosons are formed from the superpositions of these fields and are formed as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (2.12)$$

$$Z_\mu = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W, \quad (2.13)$$

and

$$A_\mu = W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W, \quad (2.14)$$

where θ_W is the weak mixing angle [5], defined by the ratio of the two electroweak coupling constants (weak isospin, g' , and weak hypercharge, g_W)

$$\tan(\theta_W) = \frac{g'}{g_W}. \quad (2.15)$$

The full EWK covariant derivative and Lagrangian density are given by

$$D_\mu = \partial_\mu + ig_W \frac{\sigma_i}{2} W_\mu^i + ig' \frac{Y_W}{2} B_\mu, \quad (2.16)$$

and

$$\mathcal{L}_{\text{EWK}} = \bar{\Psi} i \gamma^\mu D_\mu \Psi - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} W_i^{\mu\nu} W_{\mu\nu}^i, \quad (2.17)$$

respectively. If this is assumed to be an accurate representation of the EWK interaction, it requires that gauge bosons and fermions are massless in order to preserve the gauge symmetry. As it is experimentally proven that fermions and the W and Z bosons have mass, a solution is needed. The problem is solved through a process called electroweak symmetry breaking.

2.5.1 Electroweak symmetry breaking

Electroweak symmetry breaking solves the mass problem by keeping the particles massless but introducing a new complex scalar field doublet and potential into the Lagrangian density.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2.18)$$

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D^\mu \phi) - \frac{1}{2} \mu^2 \phi^\dagger \phi - \frac{1}{4} \lambda (\phi^\dagger \phi)^2 \quad (2.19)$$

This potential introduces an infinite set of degenerate non-zero ground states of ϕ . The vacuum expectation value, v , is the magnitude of an arbitrary ground state. It is non-zero if μ^2 is negative and the value inferred from measurements is:

$$v = \sqrt{-\frac{\mu^2}{\lambda}} \approx 246 \text{ GeV}. \quad (2.20)$$

The choice of v , conventionally along the neutral, real component of ϕ , spontaneously breaks the symmetry of the Lagrangian [12]. The neutral component is chosen because the charged component of ϕ must be set to zero such that the addition of a vacuum expectation value will not break the conservation of electric charge. By taking this choice, ϕ can be rewritten as a perturbation around this point

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (2.21)$$

where $H(x)$ is the massive BEH scalar field with

$$m_H = v\sqrt{2\lambda}. \quad (2.22)$$

The other three degrees of freedom from the initial scalar field doublet form the three massive weak bosons, via the BEH mechanism [13–15] such that:

$$m_W = \frac{vg_W}{2}, \quad m_Z = \frac{v\sqrt{g_W^2 + g'^2}}{2}. \quad (2.23)$$

Fermions may also be permitted mass via the Higgs mechanism, by introducing additional terms to the Lagrangian density of the form

$$\mathcal{L}_Y = g_y (\bar{\Psi}_L \phi \Psi_R + \bar{\Psi}_R \phi^\dagger \Psi_L), \quad (2.24)$$

where g_y (Yukawa coupling) is the coupling strength of the respective fermion to the BEH field. The Yukawa coupling strength is different for each fermion. This leads to fermion masses which are linearly proportional to and only dependent on g_y , given by

$$m_f = \frac{vg_y}{\sqrt{2}}. \quad (2.25)$$

2.5.2 The Cabbibo-Kobayashi-Maskawa mixing matrix

The weak interaction of quarks is described by the Cabbibo-Kobayashi-Maskawa (CKM) mixing matrix [22, 23]. It postulated the existence of a third family of quarks (t, b) to explain the observed CP violation in kaon decays [23]. It relates the weak, flavour eigenstates to the mass eigenstates by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.26)$$

The relative strength of the weak interaction involving a flavour transition from $j \rightarrow i$ is given by the associated CKM matrix element, V_{ij} . In the SM the CKM matrix is unitary ($V^\dagger V = I$) so that each row behaves as $|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1$. The current experimental measurements of the elements of the CKM matrix, given in the PDG in 2016 [12], are

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}. \quad (2.27)$$

The relative strength of the off-diagonal CKM elements means that flavour transitions between different generations of quarks are suppressed. There is a large complex phase that is present in the V_{ub} ($\approx 70\%$) and V_{td} ($\approx 20\%$) matrix elements, which allows for charge-parity (CP) violation in the SM.

2.6 The shortcomings of the SM

The full SM combines the QCD and EWK field theories into the combined symmetry group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. This has the effect of making the full covariant derivative

$$D_\mu = \partial_\mu + ig_S \frac{\lambda_a}{2} G_a + ig_W \frac{\sigma_i}{2} W_\mu^i + ig' \frac{Y_W}{2} B_\mu, \quad (2.28)$$

and the full Lagrangian density the sum of all individual components

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EWK}} + \mathcal{L}_{\text{H}} + \mathcal{L}_{\text{Y}}. \quad (2.29)$$

It includes terms for free fermions and gauge bosons and the interactions between them, in addition to mass terms arising from the introduction of the BEH field and a non-zero v . It is the most complete model we have of our Universe, describing fundamental matter and its interactions to an unprecedented degree of precision, however as mentioned earlier, it is far from a complete description.

2.6.1 Gravity

Perhaps the largest omission from the SM is the fundamental force of gravity, which all massive particles experience. In order to include a description of gravity in the SM, a unification with the theory of general relativity at the quantum scale is needed. The formulation of this unification at present is unknown, however one popular way to do this is with string theory [24]. Unfortunately, the effect of gravity is completely negligible at the current accessible energy scales and as such it is not experimentally feasible to directly test any unification theory.

2.6.2 Massive neutrinos

The SM assumes that all neutrinos are massless. Experiments such as SNO, which measured a third of the number of ν_e expected from production in the Sun [25], or Super-Kamiokande, which measured half of the number of upward atmospheric ν_μ as downward ones [26], proved that neutrinos oscillated in flavour. This oscillation can be explained by the neutrinos having a mass, such that the weak flavour eigenstates of the neutrinos are linear superpositions of the mass eigenstates, in contradiction to the SM.

2.6.3 Dark matter

Other issues with the SM originate from cosmological observations. Firstly, the lack of enough mass in galaxies to explain the rotation curves of luminous matter, provides evidence for a type of additional matter which must be massive and electromagnetically inert [27]. This matter is known as dark matter and is predicted to contribute around 80% to the matter content of the Universe. There are other cosmological indicators for dark matter including galaxy clusters [28] and the cosmic microwave background [29]. While the current SM does provide a set of particles which behave in this way, the neutrinos, there are not enough to account for all of the missing matter and they are too ‘hot’, i.e. their highly relativistic velocity is unable to explain observations.

The second issue is that the observation of the accelerating expansion of the Universe implies that there is a unknown, repulsive force present in the Universe [30]. This force is known as dark energy and accounts for 70% of the mass-energy content of the Universe. The SM has no mechanism for describing dark energy.

2.6.4 Baryon asymmetry

Another striking problem in the SM is the matter-antimatter imbalance seen in the Universe. The Universe is predominantly matter-based, however the SM predicts that matter and anti-matter are produced in equal quantities. Somehow, the antimatter present at the beginning of the Universe has been transformed or destroyed and therefore a process is required which favours matter over antimatter [31]. The SM does provide an explanation for some of this transformation in the lepton sector, from the CP-violating, complex phases of the CKM mixing matrix. The CKM matrix does not generate enough CP-violation to explain the observed matter-antimatter imbalance. It is unknown how much CP-violation is present in the neutrino mixing matrix and therefore it could provide an explanation for the baryon asymmetry.

2.6.5 Hierarchy problem

The mass of the H boson depends on quantum loop corrections in the H boson propagator. These loops originate from any massive particle of the SM. The more massive the particle in the loop, the larger the effect on the H boson mass. The mass can be factorised into its bare mass and quantum loop corrections as a function of the energy scale of the interaction, μ .

$$m_{\text{H}}^2 = m_{\text{H,bare}}^2 + \mathcal{O}(\mu^2). \quad (2.30)$$

At the EWK scale (≈ 100 GeV), these quantum loop corrections are comparable to the size of the bare mass, however as the energy scale approaches the Planck scale (10^{19} GeV), the point at which the SM is no longer valid, the corrections diverge. The very large divergences must be cancelled to keep the observed H boson mass at 125 GeV. This requires a significant amount of fine-tuning in the SM. One way this unnatural amount of fine-tuning can be alleviated, is to introduce an additional set of particles which, in part, cancel the loop corrections. SUSY is one example of a theory that creates an additional set of particles by creating a symmetry of spin in the SM [32]. For every SM particle there exists a massive supersymmetric particle, differing by a half unit of spin, which can be present in quantum loop corrections to the H boson mass at higher energy scales, cancelling contributions from the SM particles.

2.7 Top quark physics

The discovery of the top quark was announced in 1995 from a combination of the results from both the CDF (Collider Detector at Fermilab) and D0 (located in the region D0 of the Tevatron)

experiments [33, 34]. It is the most massive, elementary particle in the SM to date. The decay width of the top quark ($\Gamma_t = 1.35$ GeV) is proportional to its pole mass ($\Gamma_t \propto 1/m_t^3$) [12]. The lifetime of the top quark is the inverse of its decay width, $\tau_t = 1/\Gamma_t$, which results in a very short lifetime at $\simeq 5 \times 10^{-25}$ s [12]. In fact, it is shorter than the typical time it takes for a quark to hadronise, $\tau_{\text{had}} \approx 3 \times 10^{-24}$ s [19], so that it is possible to study the bare quark properties through its decay products.

Now, at the LHC, more top quarks are being produced in proton-proton collisions than ever before. This comes from both an increase in the luminosity and in the centre-of-mass energy, $\sqrt{s} = \sqrt{p^A + p^B}$, where p^A and p^B are the four-momenta of the colliding protons. The LHC and the CMS experiment will be discussed in detail in Ch. 3. The abundance of top quark data is leading to a vast array of measurements being performed, such as precision measurements of the SM, searches for rare SM processes and searches for possible new physics present in the production, decay and couplings of top quarks. Not only is top quark physics important in direct searches for new physics, it is often the process which forms the largest background in other searches for BSM particles. Therefore, it is crucial to know the production cross section of the top quark as precisely as possible. The precise top quark production cross section measurements can also be used to directly probe the V_{tb} element of the CKM matrix ($V_{tb} \gg V_{ts} > V_{td}$), to set constraints on the gluon and quark parton distribution functions (PDFs) of the proton, and to measure the top quark Yukawa coupling strength.

2.7.1 Top quark production at the LHC

Protons are made up of three valence quarks (uud), confined by gluons. At higher energies, the gluon multiplicity increases, as well as the number of pair-produced $q\bar{q}$ pairs, known as sea quarks. This means that it is much more probable that an interaction will occur from collisions between gluons or non-valence quarks. Top quark production at the LHC at $\sqrt{s} = 13$ TeV proceeds primarily via $t\bar{t}$ pair-production, where $\sim 90\%$ is from gluon-gluon fusion and the further $\sim 10\%$ from quark-antiquark annihilation. The quark-antiquark production mode has already been shown in Fig. 1, but is repeated, at leading order (LO) in perturbation theory, on the left panel of Fig. 6 and compared to the similar Feynman diagram for gluon-gluon fusion production mode on the right panel. The top quark can also be produced singly via the s- and

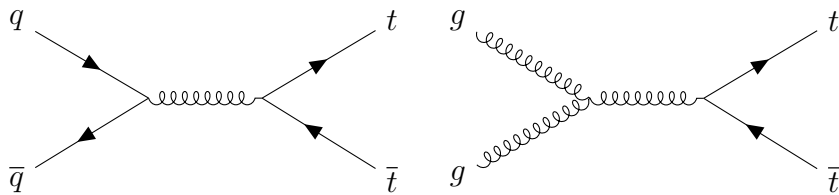


Figure 6: Leading order Feynman diagrams for the pair production of top quarks. The left panel shows pair production by $q\bar{q}$ annihilation and the right panel by gg fusion.

t- channels and in association with a W boson, as shown at LO in Fig. 7. The production cross section for single top quarks is smaller than that for $t\bar{t}$ production because it is produced only through the weak interaction. It is also smaller due to the composition of the initial states required. Higher top multiplicity final states are also available in the SM but with a much smaller production cross section [35].

The production cross section for the interaction, $\sigma_{pp \rightarrow t\bar{t}}$, is predicted by the convolution of the partonic cross section, $\hat{\sigma}_{t\bar{t}}$, and the PDFs of the incident protons, $f(x, \mu^2)$, summed over all

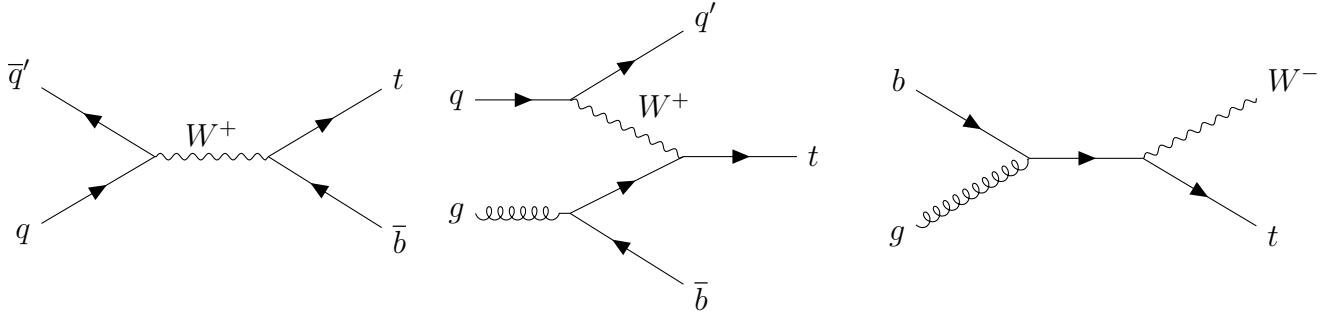


Figure 7: Leading order Feynman diagrams for the production of single top quarks. The left panel shows single top quark production via the s-channel, the central panel by the t-channel and the right panel in association with a W boson.

possible parton flavours,

$$\sigma_{pp \rightarrow t\bar{t}} = \sum_{a,b} \int_0^1 f(x_a, \mu^2) dx_a \cdot \int_0^1 f(x_b, \mu^2) dx_b \cdot \hat{\sigma}_{ab \rightarrow t\bar{t}}. \quad (2.31)$$

The energy scale at which $\sigma_{pp \rightarrow t\bar{t}}$ is valid is known as the factorisation scale.

The PDF is defined as the number density of a parton carrying a fraction x of the total momentum of the proton in the longitudinal direction at the energy scale being probed, μ . The PDFs are measured from global fits to many different data measurements and sources, such as measurements of deep inelastic scattering, Drell-Yan production, $t\bar{t}$ production and inclusive jet production, from many experiments at colliders such as the Hadron-Electron Ring Accelerator (HERA) and the LHC. The PDF is calculated at some arbitrary $\mu^2 = \mu_0^2$, which is taken as 1 GeV by the NNPDF collaboration in the NNPDF3.1 PDF set [36]. This is then evolved through the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [37–39] to higher energy scales. The $x f(x, \mu^2)$ distributions for the NNPDF3.1 PDF set at $\mu^2 = 10 \text{ GeV}^2$ and $\mu^2 = 10^4 \text{ GeV}^2$ are displayed in Fig. 8. When moving from low to high μ^2 values, it is clearly seen that a greater fraction of the momentum of the proton is carried by an increasing number of non-valence partons, primarily by gluons.

The partonic cross section of a process depends on a quantity known as the matrix-element, \mathcal{M} ,

$$\hat{\sigma}_{t\bar{t}} \propto |\mathcal{M}|^2, \quad (2.32)$$

which can be calculated directly from the Feynman diagrams. By increasing the order of perturbation theory, i.e. by adding additional partons to each diagram, makes the matrix-element calculation much more complicated but gives greatly increased accuracy. The additional partons can come in the form of virtual loops, which have been previously discussed and result in the running of the couplings, and additional, real parton emissions, such as those at next-to-leading order (NLO), shown in Fig. 9.

The prediction of the inclusive $t\bar{t}$ cross section from proton-proton collisions is shown in Fig. 10, together with current experimental measurements. The prediction from proton-antiproton collisions is also shown. The cross sections converge at higher energies due to the increase in collisions between non-valence partons, the distribution of which is identical in protons and antiprotons. At $\sqrt{s} = 13 \text{ TeV}$, the inclusive $t\bar{t}$ production cross section was calculated to be $831.8_{-29.2}^{+19.8}(\text{scale}) \pm 35.1(\text{PDF} + \alpha_S) \text{ pb}$ [40]. This $t\bar{t}$ cross section is calculated to next-to-next-to-leading-order (NNLO) accuracy in perturbative QCD, including resummation of next-to-next-to-leading-logarithmic soft-gluon terms with TOP++ (v2.0) [41–47].

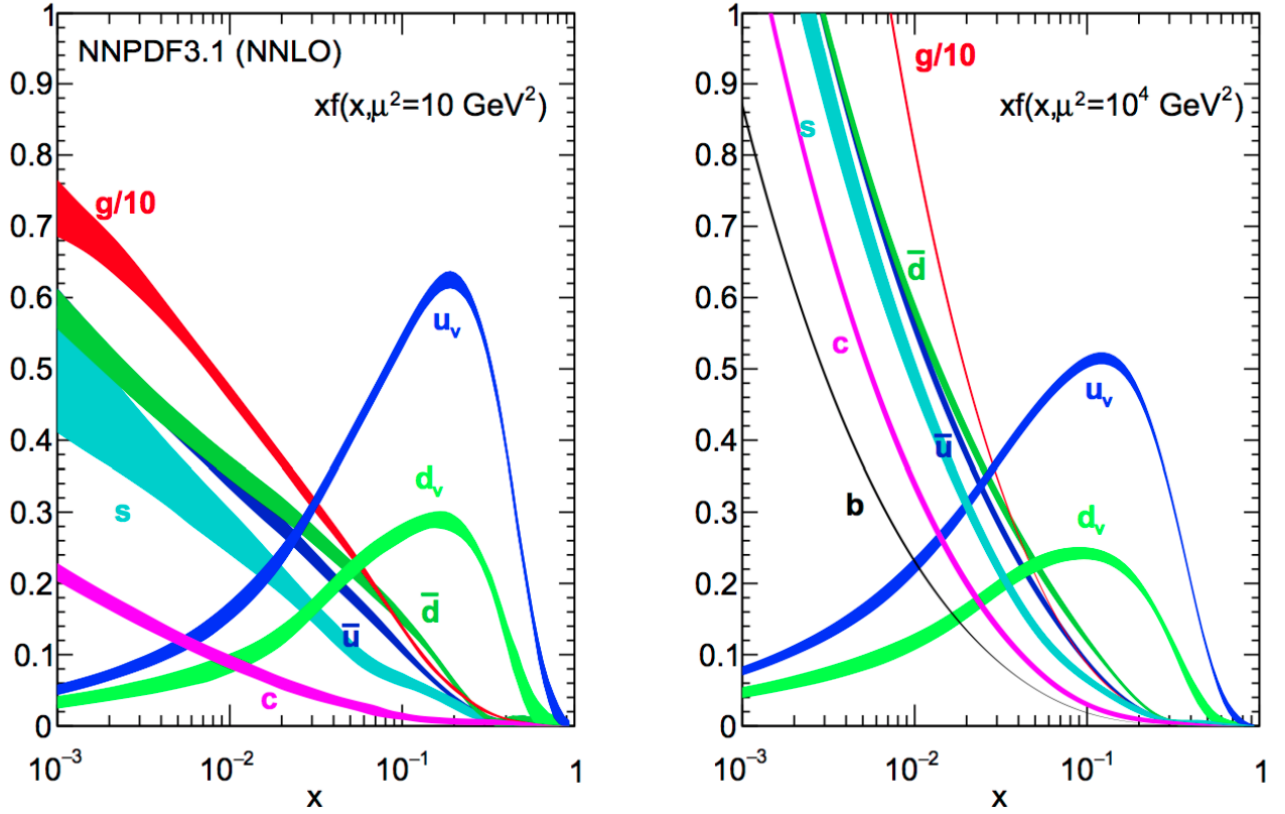


Figure 8: The parton distribution function of the proton scaled by the fraction of momentum carried by the parton, $xf(x, \mu^2)$ with respect to the fraction of momentum carried by the parton. The left panel shows the distributions with respect to a lower energy scale and the right panel with respect to a higher energy scale. Figure taken from [36].

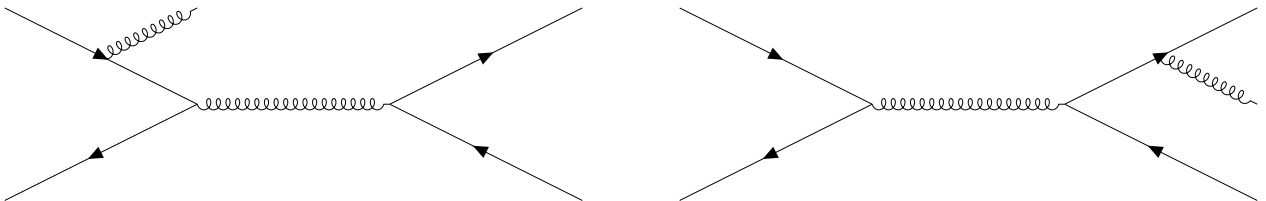


Figure 9: A set of Feynman diagrams showing real NLO processes. These are represented by the emission of real additional partons that can be seen in the final state.

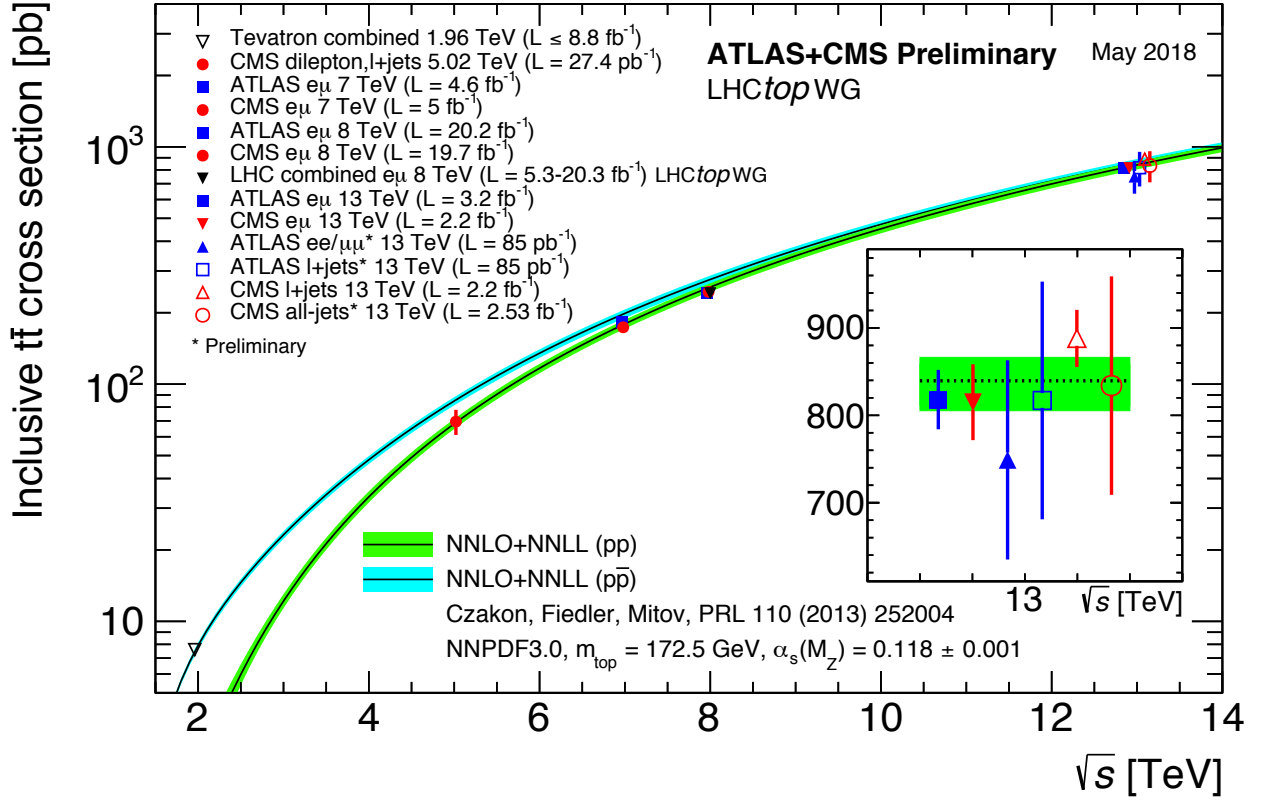


Figure 10: The inclusive $t\bar{t}$ cross section, measured by the ATLAS, CMS experiments and also at the Tevatron collider, in different decay channels at multiple \sqrt{s} . Also shown is the most precise theoretical prediction to date. Figure taken from [48].

2.7.2 Top quark decay

In the SM, the top quark will decay 99.9% [12] of the time to a b quark and W boson. The W boson then subsequently decays either into a $q'\bar{q}$ pair (hadronic) or a $\ell\bar{\nu}_\ell$ pair (leptonic). In the case of $t\bar{t}$ production, this leads to three classes of final states, the first where both W bosons decay leptonically, the second where both decay hadronically and lastly where one decays leptonically and the other hadronically. These are known as the dilepton, all-hadronic and single lepton decay channels respectively. The probability of decaying to a specific final state (branching ratio) is approximately 10% for the dilepton channel and 45% for the hadronic and single lepton channels.

Each final state category has merits and drawbacks for use in analyses. The dilepton final state provides the cleanest experimental signature; however, this comes at the cost of a reduced branching ratio. It is difficult to fully reconstruct the whole event as both neutrinos contribute to the missing transverse momentum. The all-hadronic final state, in comparison, has a large branching ratio, but comes with a large background of QCD processes. The high jet multiplicity can also cause problems from the combinatorics when reconstructing the $t\bar{t}$ system. The final state does gain in sensitivity when analysed at high collision energies. This is because the decay products of the top quark collimate into a single jet, the substructure of which can be used to identify the top quark [49]. Finally the single lepton decay channel has a large branching ratio, but with much reduced backgrounds with respect to the all-hadronic final state. The final states in which a tau lepton is produced are often neglected, however, due to the added complexity of reconstructing the tau particle. More details about the reconstruction of particles can be found in Ch. 5.

2.7.3 Top quark cross section measurements

Figure. 10 shows, in addition to the theoretical prediction of the inclusive $t\bar{t}$ production cross section, the corresponding experimental measurements, from both the CMS and ATLAS at different \sqrt{s} in all of the channels previously stated. At CMS, the inclusive $t\bar{t}$ cross section measurements are performed at 13 TeV [50, 51] at 7 and 8 TeV [52–63] and at 5 TeV [64]. Inclusive cross section measurements have been performed, by the CMS experiment, for $t\bar{t}$ production in association with a photon [65], vector boson [66] and most recently H boson [67]. The first observation of $t\bar{t}$ production in proton-lead collisions has been announced in [68]. Other inclusive cross section measurements performed by the ATLAS collaboration at 7, 8 and 13 TeV [69–79]. The Large Hadron Collider b physics experiment (LHCb), has observed the top quark for the first time in [80]. Earlier measurements of the inclusive cross section have been performed using 1.96 TeV data from the Tevatron proton-antiproton collider, the most recent of which is a combination from the D0 and CDF collaborations [81].

Measurements of the $t\bar{t}$ production cross section with respect to some other variable are known as differential cross section measurements. Differential cross section measurements are especially useful for testing theoretical models. Ideally, these models should describe $t\bar{t}$ production with respect to all kinematic variables accurately, including contributions from additional jets, and should do so for all \sqrt{s} . In practice this is not the case, with the $t\bar{t}$ models being tuned to best describe the kinematic distributions. A tune is a complete set of simulation parameters describing physics processes within the simulation of the model, which have been adapted to best describe differential data distributions. Figure. 11 shows the effect of this tuning on a set of $t\bar{t}$ models, where the new tune is shown in solid and the old in dashed, when comparing the cross sections as a function of the magnitude of the transverse momentum of the hadronically decaying top quark, $p_T(t_h)$, and the additional jet multiplicity. It is worth noting that a tune that has been applied to a specific model may end up describing the differential data not as well when applied to another independent model. Indeed by changing the tune, it may cause

the same model to describe some kinematic distributions to a worse degree, while improving other parts of the phase space. The different models behind $t\bar{t}$ production are explained in more detail in Ch. 4.

Differential top quark cross section measurements can be presented to particle level where the kinematic distributions are constructed with respect to stable particles in the detector (mean lifetime longer than 30 ps). Alternatively, they can be presented to parton level, where the results are extrapolated with respect to the final state partons or to detector level, where no extrapolation is performed. Similarly, the results can be presented in a phase space similar to that accessible by the detector, the visible phase space or extrapolated to the full phase space. Measurements that are presented to parton level or that are extrapolated to the full phase space are influenced by large theoretical uncertainties due to that extrapolation and as such this thesis presents measurements to particle level in the visible phase space. All the differential $t\bar{t}$ cross section measurements performed by the CMS experiment are complementary to each other, i.e. each measurement is performed using a unique combination of decay channels, phase spaces and final state particles such that there is little to no overlap between measurements. They are presented at 7 and 8 TeV [83–86] and at 13 TeV [3, 87–90]. Similar differential measurements have been produced by the ATLAS collaboration at these centre-of-mass energies, of which publications using 7 and 8 TeV data can be found in [91–97] and using 13 TeV data in [98–102].

This thesis looks at differential cross section measurements as a function of kinematic event variables in the single lepton decay channel (either e +jets or μ +jets) and are presented to particle level in a visible phase space. The kinematic event variables are variables that do not require the reconstruction of the complete $t\bar{t}$ system. The event variables considered are the jet multiplicity, N_{jets} , the scalar sum of the jet p_T , H_T , the scalar sum of the p_T of all particles, S_T , the magnitudes of the transverse momentum imbalance, p_T^{miss} , the p_T of the leptonically decaying W boson, p_T^W , the p_T of the lepton, p_T^ℓ , and the lepton pseudorapidity, $|\eta^\ell|$. These event variables are defined in Sec. 6.1. By measuring differential cross sections with respect to these event variables, an additional, complementary, opportunity is gained to increase our understanding of top quark production. Previous measurements with respect to kinematic event variables have been performed at 7 and 8 TeV [103] and 13 TeV using a much smaller data set [3]. The measurements in this thesis are presented in [4].

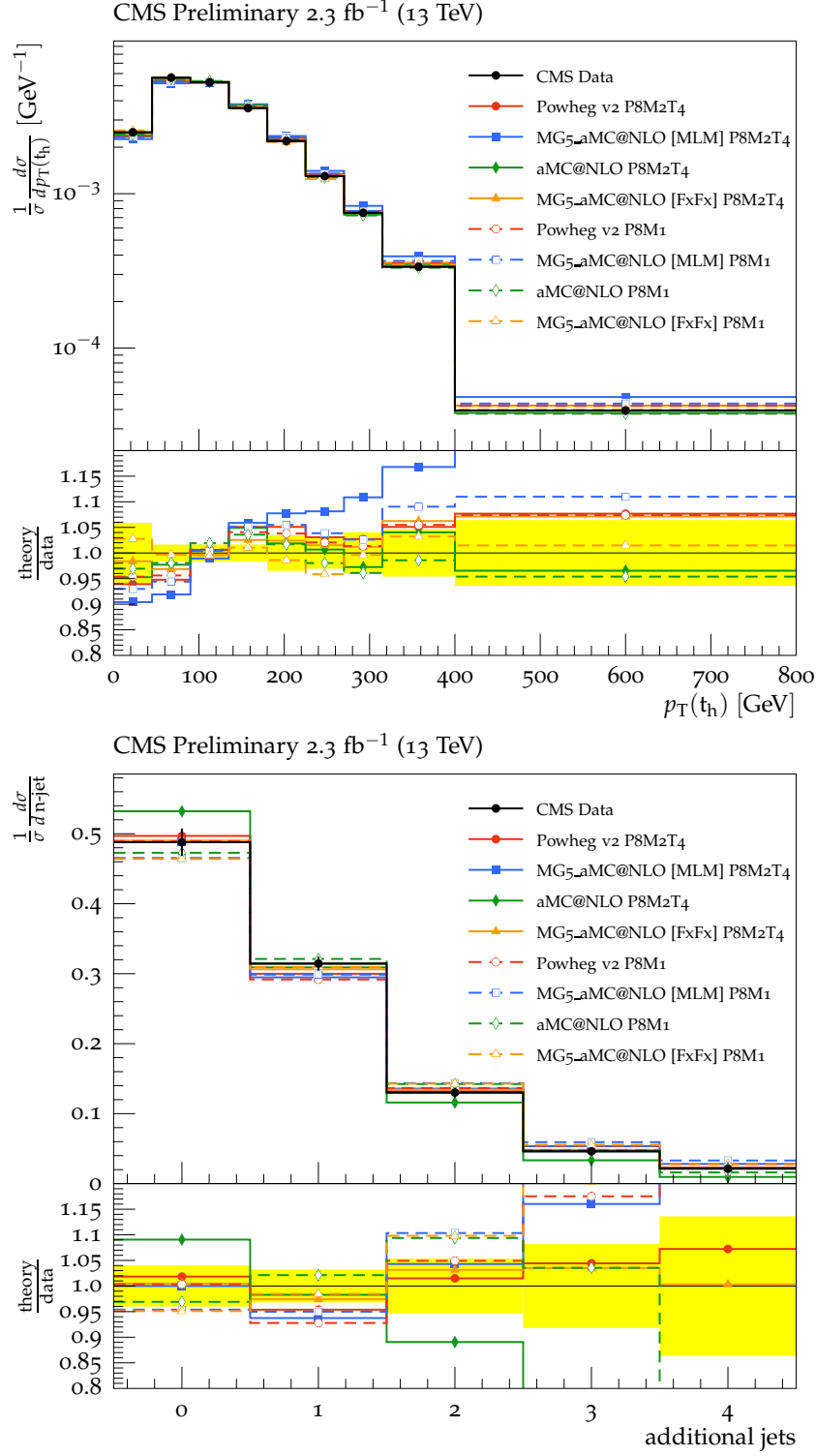


Figure 11: Comparison of different $t\bar{t}$ models with the old CUETP8M1 tune, represented by dashed lines, and the new CUETP8M2T4 tune, given by solid lines. The upper panel shows the comparison for the distribution of the p_T of the hadronically decaying top quark and the lower panel for the additional jets distribution. Figures taken from [82].

CHAPTER 3

The Large Hadron Collider and the Compact Muon Solenoid

The LHC is circular collider, situated at the Conseil Européen pour la Recherche Nucléaire (CERN) on the French-Swiss border. It is built 100 m underground in the tunnels that housed the Large Electron-Positron collider (LEP) and is currently the largest machine in the world, with an accelerator ring 27 km in circumference. For most of its operational running it collides protons with protons, however it is also able accelerate heavier ions and produce collisions such as proton-lead or lead-lead. There are four main experiments operating at the LHC: two multipurpose general detectors, CMS and ATLAS; a precision b physics experiment, LHCb; and finally a detector dedicated to heavy ion physics, A Large Ion Collider Experiment (ALICE). The full accelerator complex at CERN is shown in Fig. 12. This chapter will provide a brief history of the LHC, describe how it is operated to produce collisions in the experimental caverns and the current data-taking environment. The measurements presented in this thesis use data collected during 2016, and so a description is given of the subdetectors that formed the CMS experiment in 2016.

3.1 The Large Hadron Collider

3.1.1 A brief history of the LHC and CMS

All dates from the LHC timeline are taken from [105].

- 21st Mar 1984: A workshop was held by CERN and the European Committee for Future Accelerators (ECFA) titled ‘Large Hadron Collider in the LEP Tunnel’. It was the first official recognition of the concept of the LHC and discussed topics covering which particles should be collided and what the inherent challenges would be in the high energy collisions.
- 1st October 1992: The CMS experiment submitted its letter of intent to the LHC Experiments Committee proposing to build a multipurpose detector at the LHC. It marks the first official use of the name CMS.

CERN's Accelerator Complex

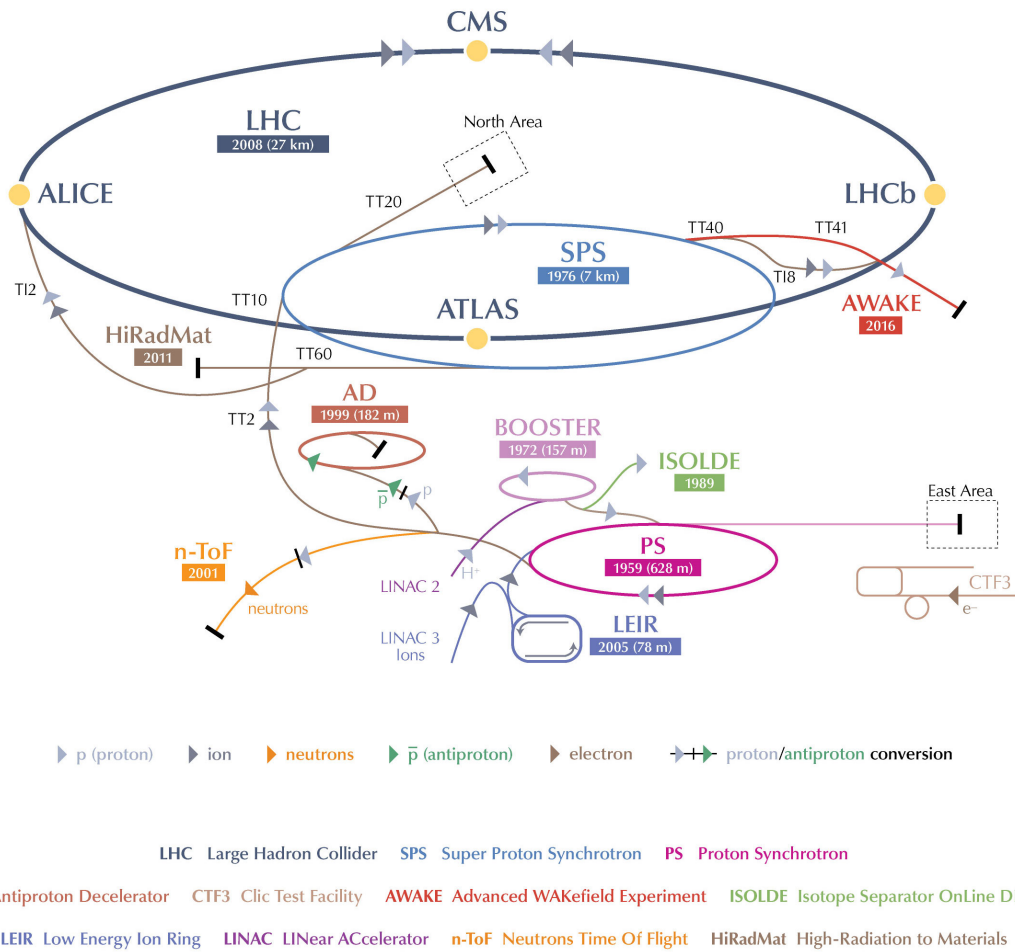


Figure 12: The full CERN complex showing schematically the location of all the experiments on the ring complex, including the location of the four main experiments at the LHC: CMS, ATLAS, LHCb and ALICE [104].

- 21st October 1993: The United States cancelled its support of the Superconducting Super Collider (SSC) project leaving the LHC as the sole candidate for a new high energy collider.
- 16th Dec 1994: The CERN council publically approves the construction of the LHC.
- 31st January 1997: The CMS experiment was approved officially as a general-purpose experiment, designed to explore the fundamental nature of the matter and interactions that shape the Universe, including searching for the H boson.
- 10th Jul 1998: Construction work began at the CMS site at Cessy.
- 1st Feb 2005: The CMS experimental cavern was inaugurated.
- 26th Apr 2007: The final superconducting magnet of the LHC was lowered into place, heralding the completion of the accelerator ring.
- 23rd Jul 2008: The final large detector segment, assembled at ground level, was lowered into the CMS experimental cavern, completing three years of detector assembly.
- 10th Sep 2008: The proton beams were first circulated around the collider, heralding a significant milestone of the LHC project.
- 19th Sep 2008: A fault occurred in an electrical connection between the superconducting magnets, leading to a release of helium in the tunnels, causing significant damage to the LHC.
- 30th Apr 2009: The final replacement superconducting magnet was lowered into the tunnels, completing repairs to the LHC.
- 20th Nov 2009: Proton beams were present in the LHC again, over a year after the incident.
- 30th Mar 2010: Run-1 physics data collection began, at $\sqrt{s} = 7$ TeV, running through until the 18th October 2011.
- 5th Apr 2012: A second period of data collection, at $\sqrt{s} = 8$ TeV, was performed lasting until the 6th Feb 2013.
- 4th Jul 2012: The discovery of the H boson was officially announced, after enough data was collected. With one of the primary mandates of the LHC fulfilled, attention has turned to discovering the existence of physics beyond the SM and measuring precisely the properties of the H boson and other SM particles.
- 3rd Jun 2015: After a two year long shutdown for upgrades, the collider started up again for Run-2 at $\sqrt{s} = 13$ TeV. Four sets of data were scheduled to be taken during Run-2, with a small set taken at the end of 2015 and major runs during 2016, 2017 and 2018. Each major run lasts from a commissioning period around March until the end-of-year technical stop in December. The measurements presented in this thesis use data collected in 2016.
- 3rd Dec 2018: The Run-2 period officially ended and now the LHC has entered a second long shutdown period, before Run-3 is scheduled to operate at $\sqrt{s} = 14$ TeV between 2020-2022.

3.1.2 Operating the LHC

The source for the proton beams is a small canister of hydrogen gas. The hydrogen gas is then ionised to create protons by applying a large electric field. The collection of protons is then accelerated, within ultra-high vacuums, through the Linear Accelerator 2 (LINAC2), which uses alternately charged conductors to accelerate the bunch of protons up to a beam energy of 50 MeV, before being injected into the Proton Synchrotron Booster (PSB). The PSB is an accelerator ring that boosts the protons to an energy of 1.4 GeV and passes them to the Proton Synchrotron (PS) which provides a further boost up to 25 GeV. From the PS the proton bunches are accelerated in the Super Proton Synchrotron (SPS) to an energy of 450 GeV. The SPS splits the single beam into two counter-rotating ones and injects them into the LHC. In the LHC, each proton beam can be accelerated up to a maximum energy of 7 TeV per beam. During Run-1, the beam energy was operated at both 3.5 and 4 TeV and was increased to 6.5 TeV for Run-2.

To reach these operational collision energies the proton bunches are accelerated, confined and squeezed by magnetic fields of up to 8.3 T from over 9300 superconducting magnets. The most common type of magnet used in the LHC is the dipole magnet, of which there are 1232. Each dipole magnet is 15 m long, weighs 35 tonnes and has a magnetic field strength of 8.3 T which is used to bend the beam around the collider. Figure 13 shows a cross section through one of the LHC dipole magnets. Each dipole contains two beam pipes for the counter-circulating beams and along each of these beam pipes are coiled niobium-titanium alloy wires which form the superconducting dipole magnet. Superconductivity in the magnet is achieved by a closed circuit of superfluid helium, which cools the iron yoke to 1.9 K.

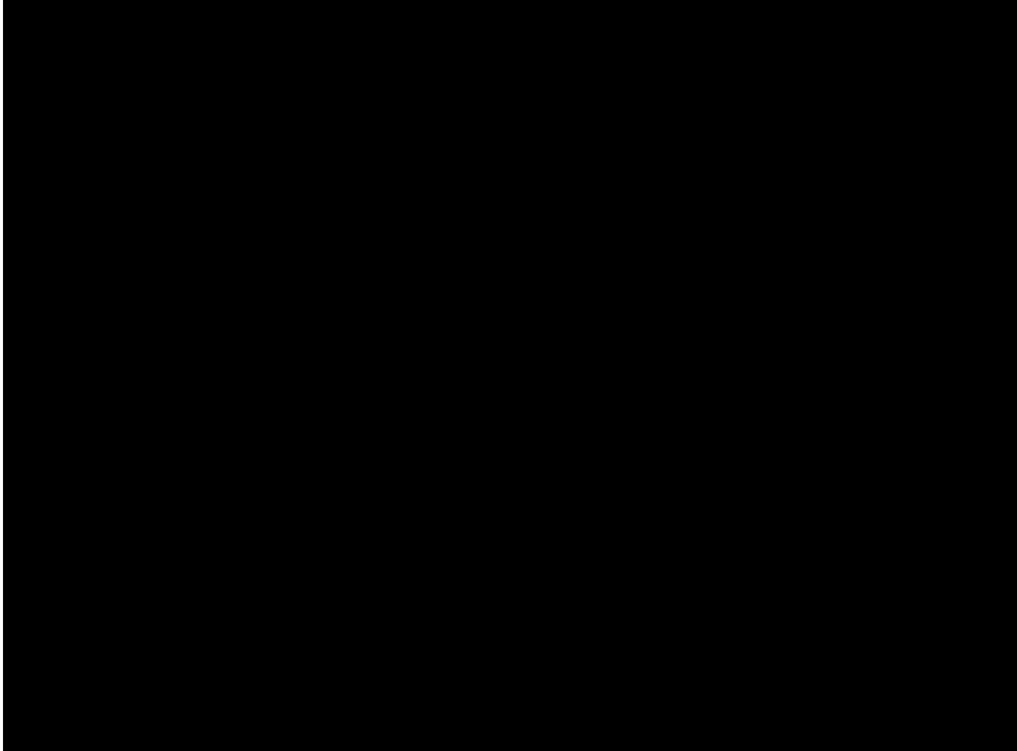


Figure 13: A cross section of a LHC dipole magnet [106].

In addition to the dipole magnets, the LHC uses 858 quadrupole magnets. Of these, 392 are lattice quadrupoles used to stop the beam from defocussing over time due to the positively charged beam constituents. Quadrupole magnets are also used to confine the proton beams as they are entering the experimental caverns. A set of three, known as an inner triplet, squeezes each beam to a diameter of $16\text{ }\mu\text{m}$ to ensure a maximal number of collisions at the interaction

point. There are additional multipole magnets used to correct for other, smaller, effects present in the beam, such as the gravitational force on the protons.

3.2 Luminosity measurements

The instantaneous luminosity, L , is a measurement of the number of collisions that can be produced by a collider in a detector, per cm^2 , per second and is given by:

$$L = f \frac{n_1 n_2}{4\pi w_x w_y}. \quad (3.1)$$

L depends on the bunch crossing rate, f , the number of protons in each colliding bunch, n_1 and n_2 , and the root-mean-square of the transverse beam widths in the horizontal, w_x , and vertical, w_y , directions. The beams are assumed to have a Gaussian profile and to be colliding head-on, with the beam widths measured to be $w = 16 \mu\text{m}$. For nominal running during Run-2, each beam contains 3564 bunch spaces of which typically 2808 are filled during data taking, leading to a minimum bunch spacing of 25 ns and a collision rate $f = 40 \text{ MHz}$. Each bunch contains $\approx 1.1 \times 10^{11}$ protons, producing a peak $L = 1.50 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. L varies with respect to time due to the depletion of the protons in each bunch after every collision. The decrease in luminosity is mitigated to a small extent by the reduction of the beam widths w_x and w_y , however each beam has an approximate physics taking lifetime of ten hours before they need to be refilled. The total number of inelastic collisions (known as minimum bias events) produced at the LHC, over a time t , can be estimated from the luminosity by:

$$N = \sigma_{\text{minbias}} \times L_{\text{int}}, \quad (3.2)$$

where σ_{minbias} is the minimum bias cross section. L_{int} is the integrated instantaneous luminosity over time t

$$L_{\text{int}} = \int_{t_0}^t L(t) dt. \quad (3.3)$$

Figure 14 shows the L_{int} delivered by the LHC, recorded by the CMS experiment, and validated as good for physics, during data taking in 2016. The validation requires good performance in both the detectors and reconstructed physics objects. A total of 35.9 fb^{-1} of data was certified to be used for physics. The integrated luminosities delivered by the LHC across the data taking years of Run-1 and Run-2 are shown in Fig. 15 and displays the exceptional performance of the machine. Using the current σ_{minbias} measurement of 69.2 mb, this means that approximately 2.5×10^{15} collisions were processed in this period. More than one proton-proton interaction typically occurs in each bunch crossing and these additional interactions are referred to as in-time pileup. As the collision rate at the LHC is so high, it is also possible for remnants of previous bunch crossings to affect the current one, particularly in calorimeters where there is a latency higher than a typical bunch spacing. These remnants are called out-of-time pileup. The average number of interactions per bunch crossing during 2016 data taking was approximately 25, as can be seen in Fig. 16, using a minimum bias cross section of 69.2 mb [109]. Particles from pileup are included in the particle reconstruction of the event, leading to misidentification. Reconstruction and pileup mitigation techniques are described in Secs. 5.1.5 and 5.2.3.

CMS Integrated Luminosity, pp, 2016, $\sqrt{s} = 13$ TeV

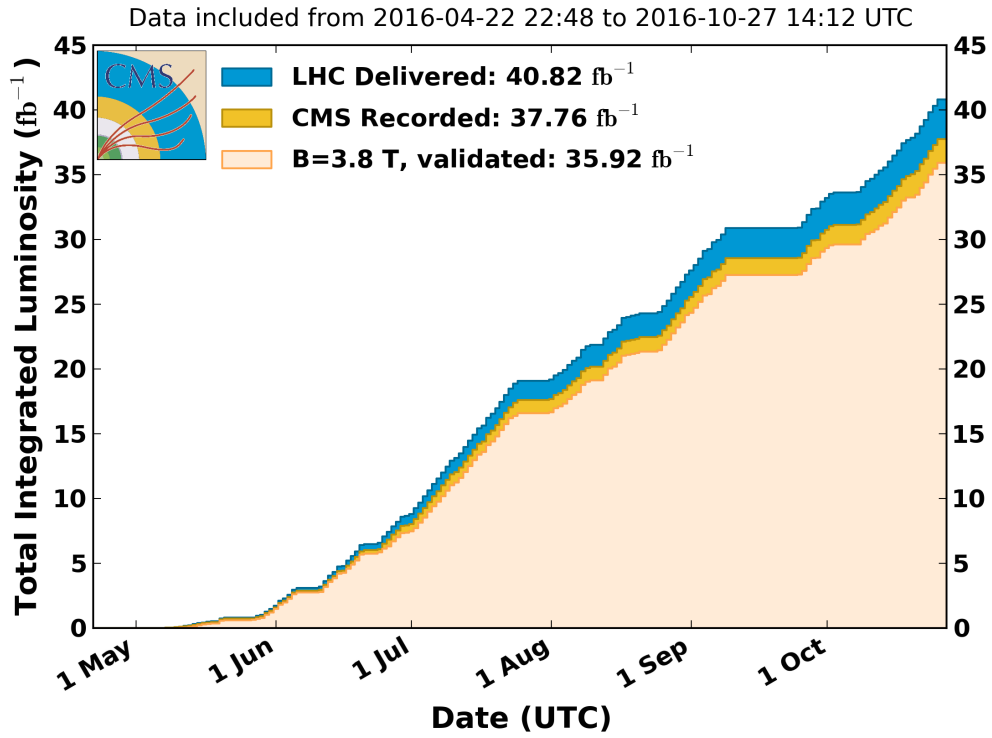


Figure 14: The integrated luminosities delivered by the LHC and recorded by the CMS experiment during data taking in 2016. Figure taken from [107].

CMS Integrated Luminosity, pp

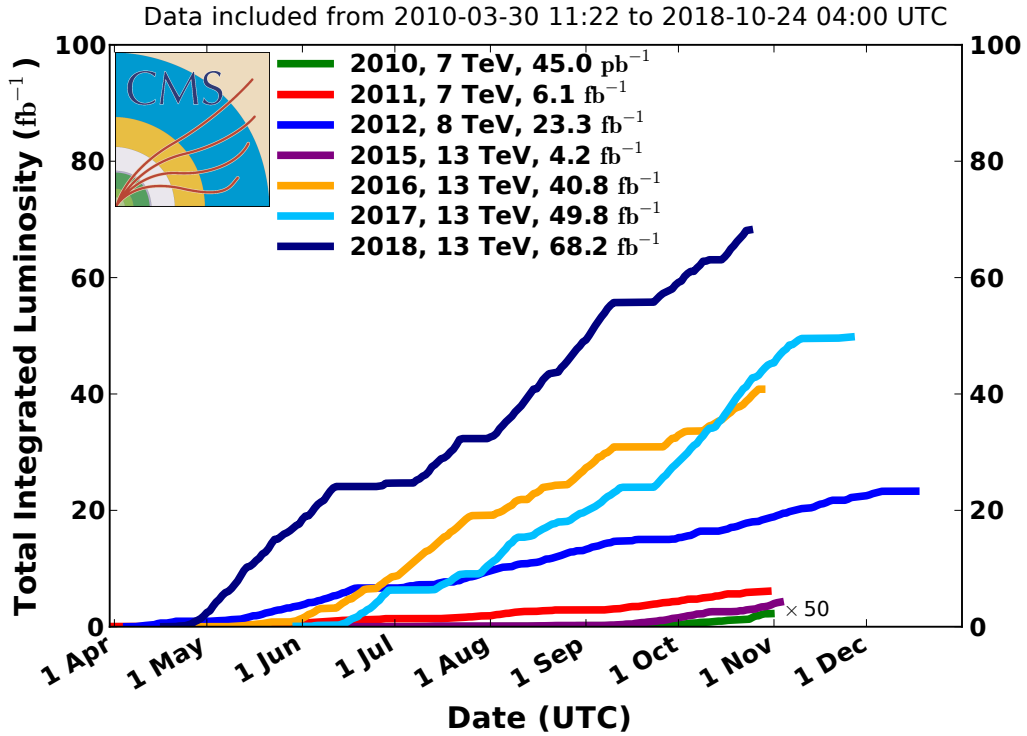


Figure 15: The integrated luminosity delivered by the LHC over all years of operation. Figure taken from [108].

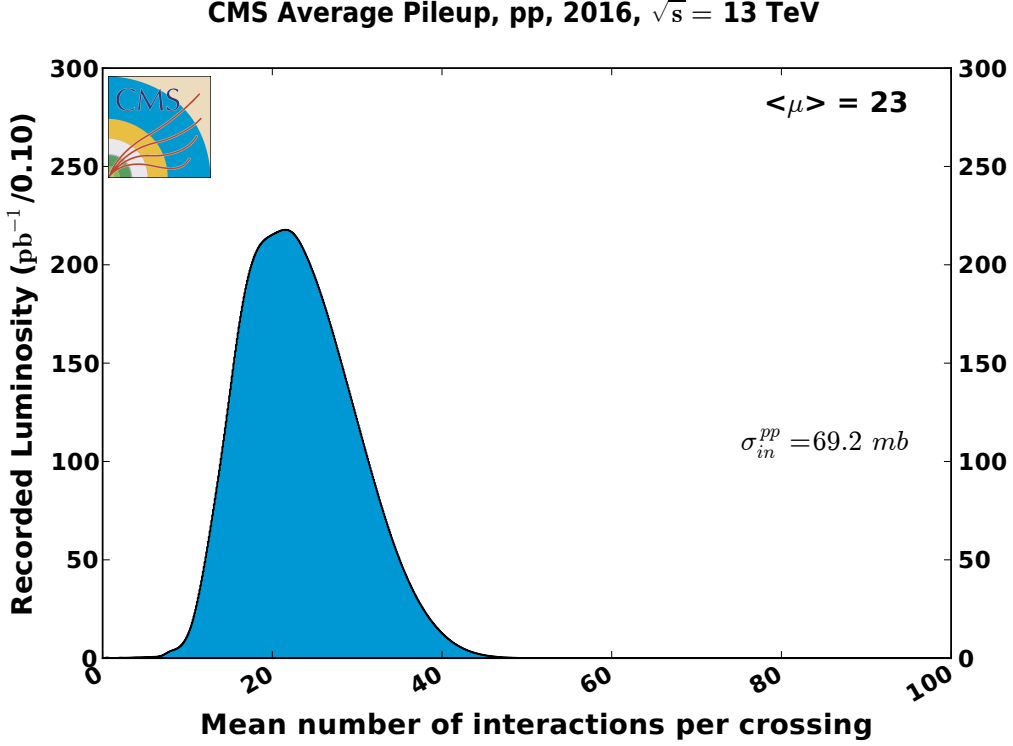


Figure 16: The distribution of the number of interactions in a bunch crossing. The minimum bias cross section is taken to be 69.2 mb. Figure taken from [108].

3.3 The Compact Muon Solenoid

CMS is a general purpose detector, designed to be able to perform precision measurements of the SM and searches for new physics [10]. It is approximately 15 m in diameter and 30 m long, with a mass of 14 000 tonnes. This high density gives the experiment its name ‘compact’, being almost a quarter of the volume and double the mass of its sister experiment ATLAS. It is a cylindrical, hermetic detector composed of a barrel region and two endcaps, each formed of layers of subdetectors around the beam collision point, as shown in Fig. 17. These subdetectors, in order from the collision point, are the silicon pixel and strip trackers, electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL) and muon chambers. The superconducting magnet is situated in between the hadronic calorimeter and muon chambers. To operate in a high luminosity environment the CMS experiment must have excellent resolution in each of its subdetectors and a low latency to manage the high collision rate. The materials used must be reliable while having maximal longevity when operating in a high-radiation environment.

CMS uses a right-handed coordinate system, where the x -axis points towards the centre of the accelerator ring, the y -axis points vertically upward and the positive z -axis lies parallel to the anticlockwise beam axis. The azimuthal angle, ϕ , is measured from the x -axis in the transverse $x - y$ plane. The polar angle, θ , is measured from the positive z -axis in the $z - y$ plane. At the CMS experiment, collisions are viewed in the centre-of-mass frame of the colliding protons, which means that the colliding partons usually have a Lorentz boost along the beam direction. For this reason, the angles of the particles are normally expressed in terms of the rapidity, y , or pseudorapidity, η , in which the production is roughly constant. The rapidity is defined as

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (3.4)$$

where E is the energy of a measured particle and p_z is the z -component of the momentum.

Rapidity is advantageous to use as differences in the rapidity between two particles are invariant under a Lorentz boost along the z -axis. At the LHC, most particles are highly relativistic such that the rest mass is small compared to the energy ($p_z \approx E \cos \theta$). This can be substituted into Eq. 3.4 and, using trigonometric relations, the pseudorapidity is defined as

$$\eta \equiv -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (3.5)$$

Jet production in proton-proton collisions is roughly constant in η , reflecting the forward nature of the production. The subdetectors at the CMS experiment extend to at least a pseudorapidity magnitude of 2.4, covering more than 95% of the angular phase space. Individual subdetectors can extend to a higher $|\eta|$, for example the forward detectors of the HCAL, to study very boosted jets or the underlying event. The acceptance within $|\eta| < 2.4$ can be hampered by issues such as the gap between barrel and endcap detectors.

To completely describe the kinematic properties of a particle, its energy and momentum must be known. While it is not possible to detect the complete momenta spectrum of an event, due to many particles produced in the collision being lost down the beam pipe and the unknown boost of the initial state partons, it is possible to measure the momenta in the transverse plane where the sum of the momenta is conserved. To measure the transverse momenta and energy of particles CMS uses a combination of hits in the tracking systems from charged particles and energy deposits in the calorimeters. These subsystems are described in more detail in the following sections.

3.3.1 Silicon pixel and strip trackers

Figure 18 shows the layout of the silicon tracking detectors in CMS used during data taking in 2016. Closest to the interaction point are the pixel detectors, consisting of three barrel layers situated at radial distance $r = 4.4, 7.3$ and 10.2 cm in the transverse plane from the interaction point, with two endcap disks lying at $z = \pm 34.5$ and ± 46.5 cm. There are 65 million pixels in the inner tracker, each $100 \times 150 \mu\text{m}^2$, giving a total active area of 1.06 m^2 . Surrounding the pixel detectors are the silicon strip tracker sensors. These are arranged into four subsystems: the Tracker Inner Barrel and Disks (TIB and TIDs), the Tracker Outer Barrel (TOB) and the Tracker Endcaps (TECs). The TIB consists of four barrel layers between $20 < r < 55$ cm with each end capped by 3 disks ($|z| < 118$ cm) forming the TIDs. Wrapping around the TIB and TIDs is the TOB consisting of six barrel layers between $55 < r < 116$ cm. Finally, each TEC consists of nine disks between $124 < |z| < 282$ cm, capping the TIB, TIDs and TOB. There are 15 148 silicon strip modules in total that make up the outer tracker, giving a total active surface area of 198 m^2 . Some strip layers in the tracker, coloured blue in Fig. 18, have back-to-back modules, rotated with respect to each other by a small stereo angle ($\approx 5^\circ$) in order to improve the 3-dimensional point resolution by providing a measurement of the z co-ordinate in the barrel and r in the disks.

Each tracking sensor operates in a similar manner. For the pixel sensors, an $n^+ - n$ junction is utilised where the silicon has been doped with phosphorous to different degrees. For the strip sensors, a $p^+ - n$ junction is used where a silicon strip doped with boron has been laid on the bulk n -type silicon. Doping with phosphorous causes the silicon to become an electron donor (n -type) and doping with boron to become an electron acceptor (p -type). The $n(p)^+$ -type silicon is where silicon has been doped to such an extent that the resistivity is very low.

In the case of the silicon strip sensor, electrons diffuse from the bulk n -type silicon to a strip of p^+ -type silicon, creating a small opposing electric field from the formation of ions either side of the $p^+ - n$ junction. The field stops further electron-hole movement and forms a potential step. The region with no excess electrons or holes is known as the depletion zone, which causes

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel ($100 \times 150 \mu\text{m}$) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying $\sim 18,000\text{A}$

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

FORWARD CALORIMETER
Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator $\sim 7,000$ channels

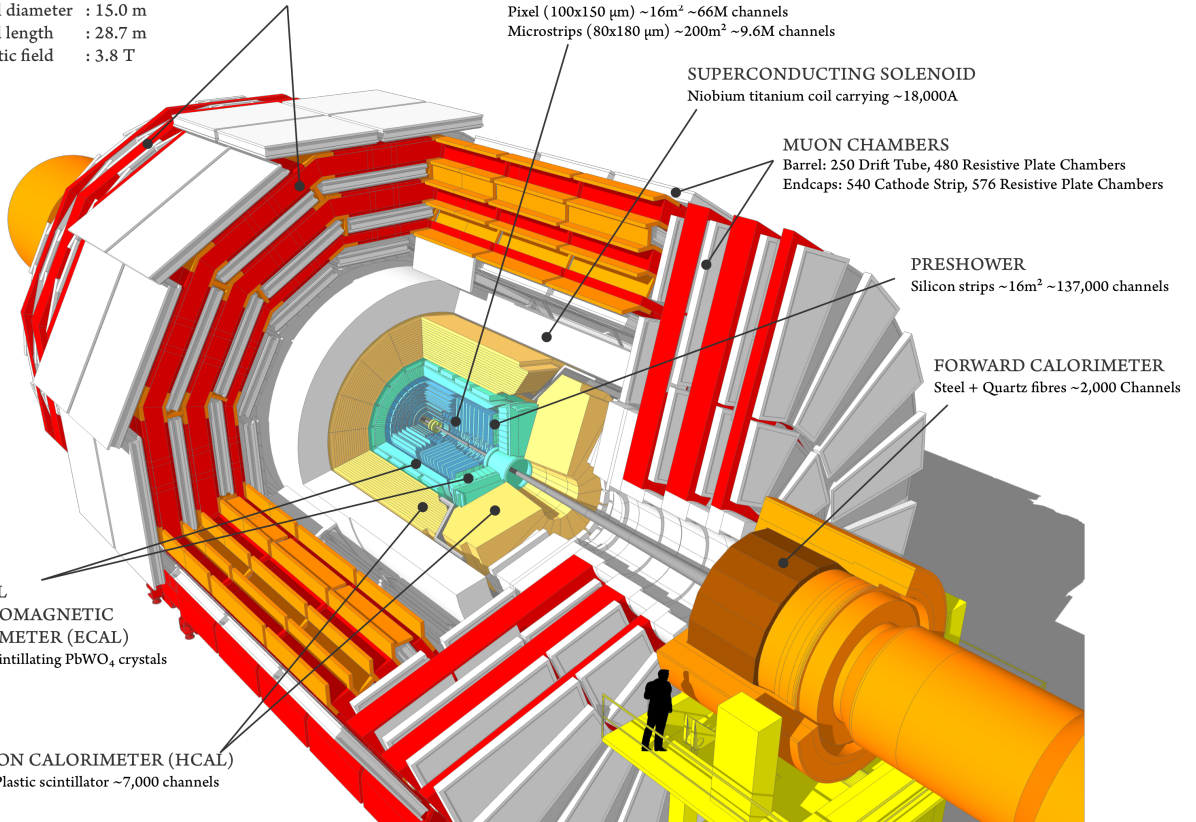


Figure 17: A schematic of CMS as used in 2016, sliced to show the internal layout. It consists of layers around the central interaction point in the order of silicon pixel and strip detectors, electromagnetic and hadronic calorimeters, superconducting magnet and finally muon chambers [110].

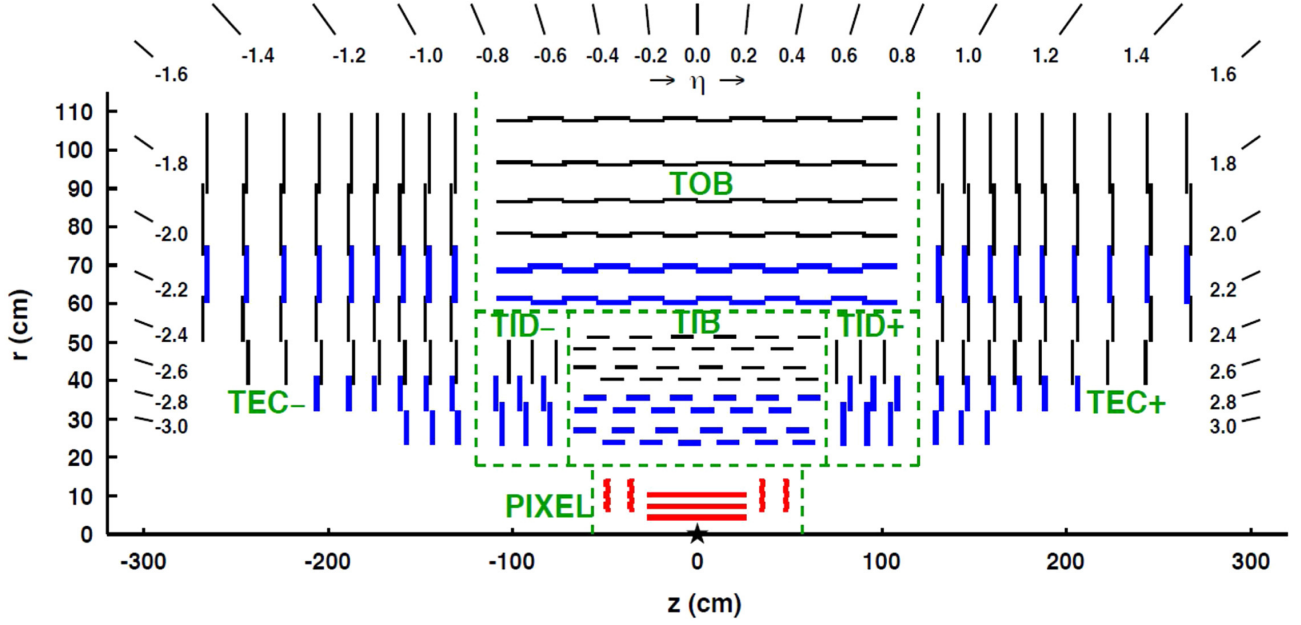


Figure 18: A slice through the CMS tracker detector in the $r - z$ plane. The separate subsystem regions have been highlighted and are classified as the Tracker Inner Barrel and Disks (TIB and TIDs), the Tracker Outer Barrel (TOB) and the Tracker Endcaps (TECs). The Pixel subsystem is highlighted in red. Back-to-back stereo strip modules are highlighted in blue. Figure taken from [111].

the junction to act as a diode. A reverse bias voltage is applied to ensure the silicon is fully depleted by attracting the free electrons and holes away from the $p^+ - n$ junction. When a charged particle ionises the n-type bulk silicon, the electrons and holes produced ($\approx 30\,000$) drift in the large electric field such that the holes move towards the p^+ -type silicon producing a small signal current which is then amplified and shaped. This method can be applied to any semiconductor junction where each side has different doping concentrations.

The pixel detector receives a charged particle flux of 1 MHz mm^{-2} . At this high rate of particles the pixels must have a high granularity to keep the fraction of channels hit (occupancy) reasonable. A low occupancy in the pixel detector is necessary to reconstruct individual particle tracks back to an interaction vertex, in a high-track environment. Lower occupancies present in the TOB and TECs mean that a longer strip can be used, however this incurs an increase in noise, which is alleviated by increasing the thickness of the silicon sensor from $320\text{ }\mu\text{m}$ to $500\text{ }\mu\text{m}$. Tracks reconstructed in the pixel and strip detectors can then be used in vertex reconstruction, particle identification and charged particle momenta measurements. The track reconstruction is detailed in Sec. 5.1.1.

3.3.2 Electromagnetic calorimeter

The CMS ECAL, as shown in Fig. 19, is made of three subsystems: the barrel, the endcaps and the preshower detector. Together, they form a compact coverage around the interaction point, up to $|\eta| < 3$. A total of 75 848 lead tungstate (PbWO_4) scintillating crystals are used in the ECAL barrel and endcaps. In the barrel section, the crystals are laid out in an $\eta - \phi$ grid extending to $|\eta| < 1.479$. The inner surface of the barrel is at $r = 1.29\text{ m}$, with each crystal having a front surface area of $2.2 \times 2.2\text{ cm}^2$ and length 23 cm. In each endcap, the crystals are instead laid out in an $x - y$ grid, with the front surfaces at $z = \pm 3.14\text{ m}$, extending through $1.479 < |\eta| < 3.0$. Each endcap crystal has a front surface area of $2.9 \times 2.9\text{ cm}^2$ and a length of 22 cm. All the crystals are slightly off-centre with respect to the primary interaction point

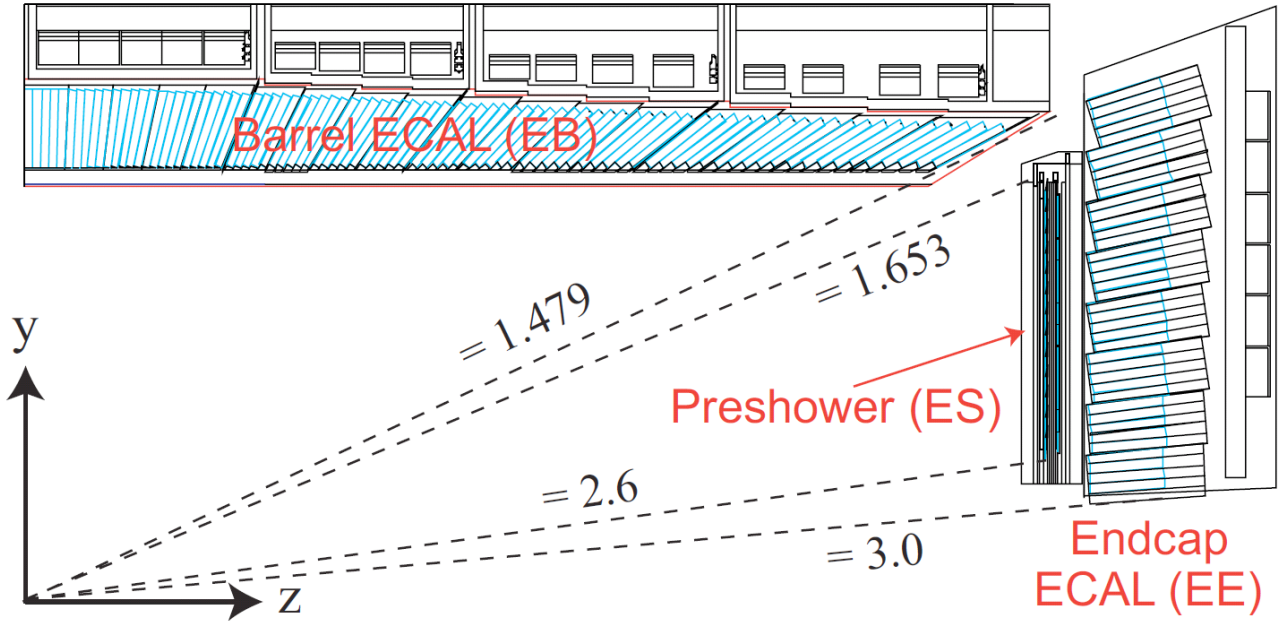


Figure 19: A geometrical quarter view of the CMS ECAL, highlighting the barrel, endcap and preshower detector regions. Figure taken from [112].

to avoid particles being lost down the small gaps between adjacent crystals.

PbWO_4 is an ideal choice for the CMS ECAL due to its short radiation and Molière lengths of 0.89 and 2.2 cm respectively. The radiation length gives a measure of the penetration of the EM shower into the crystal and the Molière length how confined the EM shower is, such that a cascade will mostly be contained within one crystal. In order to contain $> 99.9\%$ of the EM shower, the length of the calorimeter crystals must be at least 25 times that of the radiation length and as wide as the Molière length. In addition, PbWO_4 crystals are very fast, with 80% of the blue-green scintillation light being emitted in the time it takes for another collision to occur. The light emission from the PbWO_4 crystals is relatively small at 30γ per MeV of energy deposited and therefore needs to be amplified using silicon avalanche photodiodes (barrel) and vacuum phototriodes (endcaps). PbWO_4 is also radiation resistant tolerating up to $6.2 \times 10^{17} \text{ MeV kg}^{-1}$ of radiation energy.

The purpose of the ECAL preshower detector is to distinguish between a high energy single photon and a π^0 decaying into two closely spaced photons. To do this the preshower consists of two layers of lead, which initiate an EM shower for photons (and electrons), each closely followed by a layer of silicon strip detectors. The silicon strips are placed orthogonal to each other to measure shower positions precisely and are fine enough in granularity to differentiate between a single photon shower and multiple close-spaced photons. A π^0 decaying in the barrel typically has low enough energy not to need the additional resolution provided by the preshower detector. The preshower degrades the resolution of the ECAL endcap due to the dense lead plates, however as the energy deposited within the lead is proportional to that deposited in the silicon, a correction can be applied.

These attributes mean that the ECAL can be both compact enough to fit within the magnet and HCAL and granular enough to perform with high energy resolution. The resolution performance is measured by fitting a Gaussian function to the reconstructed energy distributions at different test beam energies and is parametrised by

$$\frac{\sigma_{\text{ECAL}}}{E} = \frac{S}{\sqrt{E}} + \frac{N}{E} + C, \quad (3.6)$$

where S is a stochastic term (number of photoelectrons), N is a noise term (electronics and

digitisation) and C is a constant term (leakage from the back of the ECAL). Figure 20 shows the energy resolution with respect to the energy deposited in a 3×3 crystal lattice around incident electrons measured under test beam conditions before the beginning of Run-1 with the values for S , N and C shown. Figure 21 shows the energy resolution with respect to $|\eta|$ from the most recent public measurement [113]. It is extracted from an unbinned likelihood fit on $Z \rightarrow e^+e^-$ events collected in 2015 for electrons with low bremsstrahlung radiation and $E_T \approx 45$ GeV. The resolution was measured to be $< 2\%$ for $|\eta| < 1.0$ and between $2 - 5\%$ elsewhere, consistent with similar measurements during Run-1, however the lower resolution at high $|\eta|$ is caused by the limited amount of data available in this measurement.

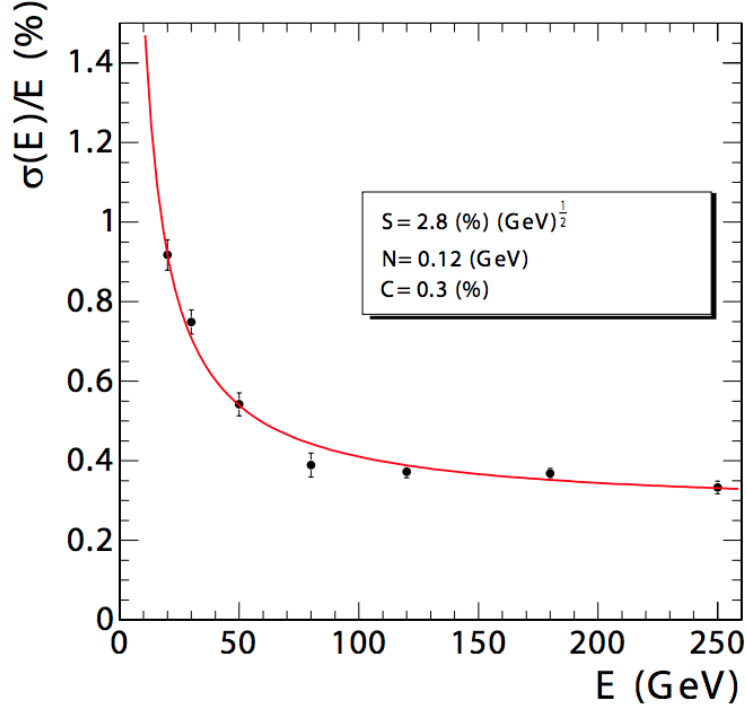


Figure 20: The energy resolution of the CMS ECAL measured in a beam test. The energy was measured in a 3×3 crystal lattice with the electron striking the central crystal. Figure taken from [10].

3.3.3 Hadronic calorimeter

The majority of the hadronic calorimeter, shown in Fig. 22, sits within the superconducting solenoid. It contains sampling calorimeters in the barrel (HB) and in the two endcaps (HE). The HB operates to $|\eta| < 1.4$ and the HE covers the range to $1.4 < |\eta| < 3.0$. The HB is radially restricted between the ECAL and the superconducting magnet at $1.77 \text{ m} < r < 2.95 \text{ m}$ and as such there is only a limited amount of absorber present necessitating the requirement for scintillator plates situated outside the solenoid (HO). The HB is divided into two half-barrels each composed of 18 identical azimuthal wedges consisting of alternating layers of non-magnetic, short radiation-length brass absorber and plastic scintillator tiles. Each wedge is capped by steel absorber for structural strength. The HE are formed of a similar structure. Around 70 000 scintillating tiles of dimensions $\Delta\eta \times \Delta\Phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$ are used in the construction of the HCAL.

The forward hadronic detectors (HF) are situated $\pm 11.2 \text{ m}$ from the primary interaction point within a pseudorapidity range of $3.0 < |\eta| < 5.2$. The HF, which is very close to the beam line, receives a large flux of particles and therefore must be very radiation resistant. To

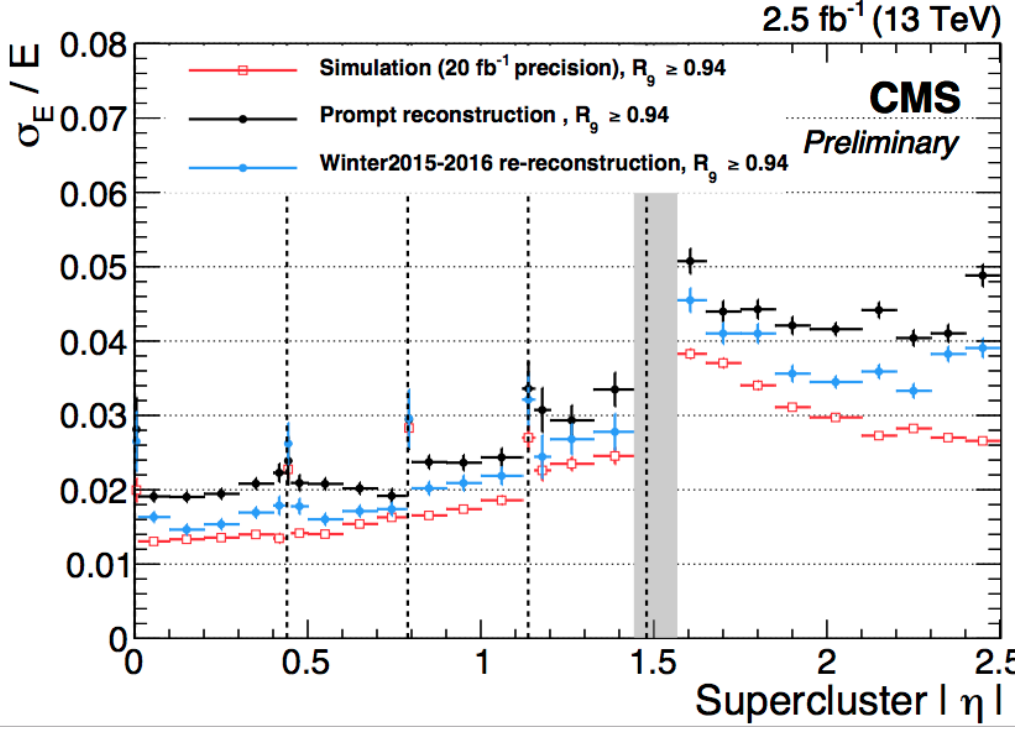


Figure 21: The energy resolution of the CMS ECAL extracted from an unbinned likelihood fit on $Z \rightarrow e^+e^-$ events. Figure taken from [113].

this end it is made of steel and quartz fibres, where the emitted Cherenkov radiation in the quartz fibres is passed to photomultipliers and read as signal [114]. The HF allows for the reconstruction of forward jets and a better description of p_T^{miss} .

As hadrons pass through the ECAL they leave only minimal deposits before entering the HCAL. When they continue through the dense absorption layers they interact and produce cascades of particles. These particle showers cause the plastic scintillator to emit blue-violet light which is passed by very small wavelength-shifting fibres as green light to the readout where the signal is amplified using hybrid photodiodes. At small $|\eta|$ the stopping power of the HB is not enough to fully contain the hadronic cascade so the HO is used to catch the tails of the hadronic cascades, improving the p_T^{miss} resolution of the detector. The resolution of the HCAL is parametrised as

$$\frac{\sigma_{\text{HCAL}}}{E} = \frac{S}{\sqrt{E}} + C, \quad (3.7)$$

where S is measured to be 0.85 and C to be 0.07 in [115]. This is similar to the expected resolution of hadronic cascades shown in Fig. 23 which are given in terms of reconstructed jet transverse energy. Jet clustering will be discussed in Sec. 5.1.5.

3.3.4 Superconducting magnet

The solenoid at the heart of CMS is the largest superconducting magnet that has ever been built at 13 m long and 6 m in diameter. It is composed from coils of a niobium-titanium alloy wire that are cooled to an operating temperature of 4.5 K. It operates at a field strength of 3.8 T, reduced from a design operation field strength of 4 T in order to maximise longevity. The magnet and steel return yokes, which are used to contain the magnetic field, are the most massive components in CMS weighing approximately 12 000 tonnes. The high field strength allows for very precise measurements of the curvature of charged-particle tracks leading to precise measurements of the momenta of the associated charged particles.

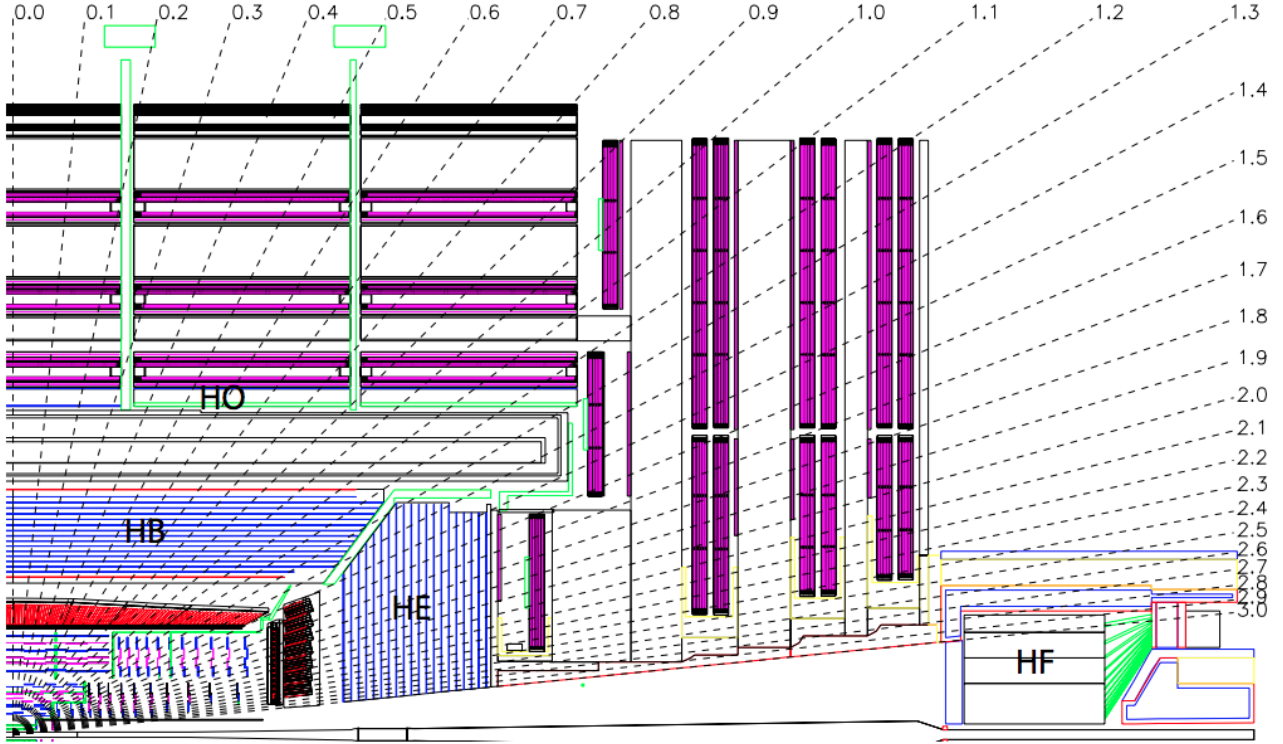


Figure 22: A geometrical quarter view of the CMS HCAL, highlighting the Barrel (HB), Endcap (HE), Forward (HF) and Outer (HO) subdetectors. Figure taken from [10].

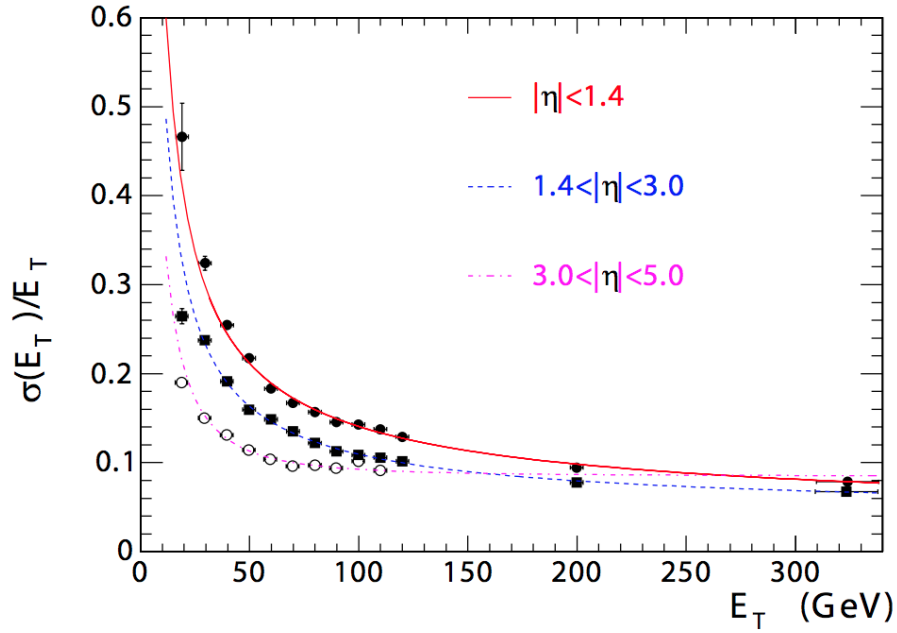


Figure 23: The transverse energy resolution of reconstructed jets in the CMS HCAL. It is split into the barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$) and forward jets ($3.0 < |\eta| < 5.0$). Figure taken from [10].

3.3.5 Muon detectors

The muon chambers are the outermost set of detectors at CMS, interleaved with the steel return yoke of the superconducting solenoid. There are three types of muon chamber present at CMS: resistive plate chambers (RPCs), cathode strip chambers (CSCs) and drift tubes (DTs). The layout of the muon chambers is presented in Fig. 24.

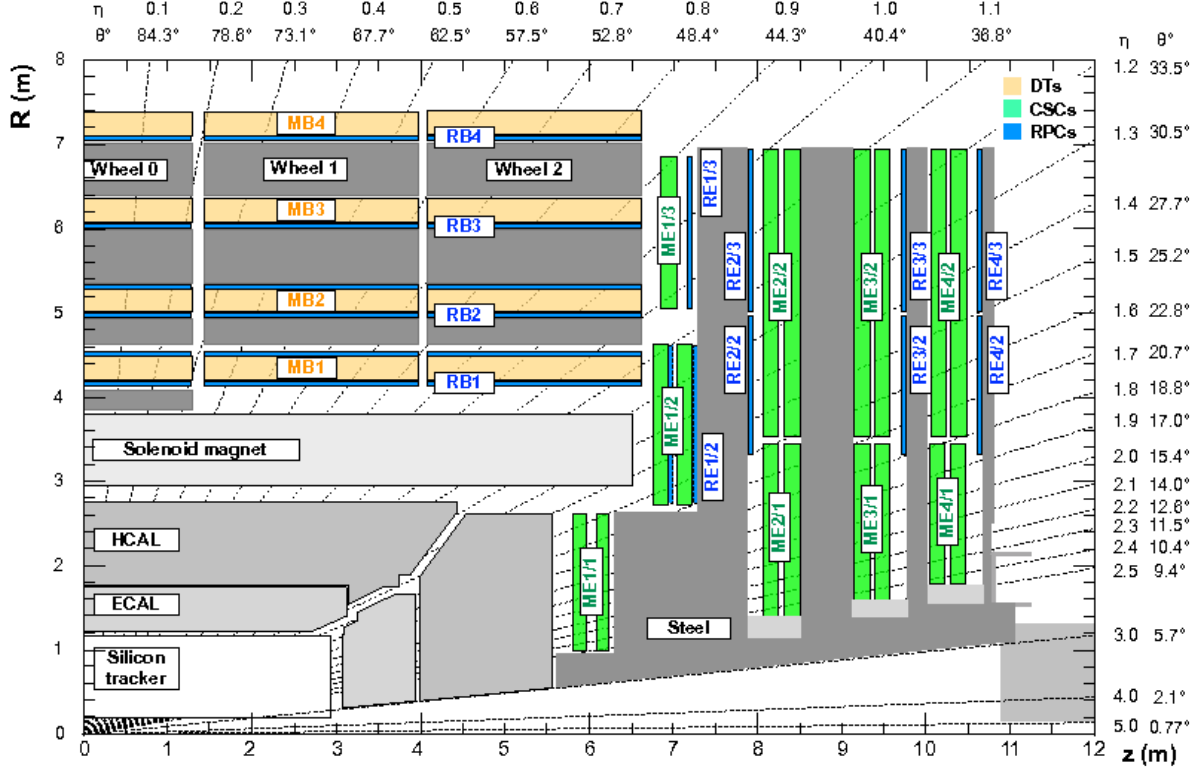


Figure 24: A geometrical one-quarter view of the CMS muon subsystems. The drift tube stations are indicated in yellow the cathode strip chambers in green and the resistive plate chambers in blue. Figure taken from [116].

The relatively low muon rates and uniform local magnetic field strengths in the barrel make DT stations a good option to measure muons up to $|\eta| < 1.2$. Each cell is $13 \times 42 \text{ mm}^2$ containing a gaseous mixture composed of 15% CO_2 and 85% Ar, with an anode central wire running the length of the cell. As a charge is left in the cell, liberated electrons drift towards the central wire forming a signal. The cells are arranged into a set of four layers, each offset by half a cell-width to reduce non-sensitive detector regions, in what is known as a superlayer. A set of three of the superlayers form a DT muon station, where the outer two are aligned parallel to the beam line and the central one perpendicular. This means the central superlayer can be used to calculate the z co-ordinate and the outer two to measure the co-ordinates in r, ϕ space. The DT cells have a high spatial resolution but also a slow response time of up to 380 ns.

In the endcaps where the muon rates are higher, 540 CSCs are employed in the range $0.9 < |\eta| < 2.4$. They have trapezoidal shapes to provide contiguous ϕ coverage in four disks in each endcap. The CSCs have a quicker response time, a finer granularity and are less affected by variations in the magnetic field making them preferable to the DTs. Each CSC station consists of seven layers of cathode panels, each with strips milled running lengthwise at constant $\Delta\phi$ width, interleaved with six layers of azimuthally running anode wires. The formation of six gas gaps are filled with 40% Ar, 50% CO_2 and 10% CF_4 . As a charged muon passes through the gas, electrons travel to the anode wires and ions to the cathode strips, providing the co-ordinates in r, ϕ space.

Finally, 610 RPCs are found in both the barrel (six layers) and endcaps (four layers) up to $|\eta| < 1.8$. They are composed of gaseous parallel-plate detectors with two gas gaps filled with a mixture of 96.2% $\text{C}_2\text{H}_2\text{F}_4$, 3.5% iso- C_4H_{10} and 0.3% SF_4 , separated by a common charge read-out strips. The RPCs have adequate spatial resolution, however they are complimentary to the DTs and CSCs due to their very fast response time (≈ 2 ns) which is much faster than the time it takes for the next bunch crossing to occur. This makes the RPCs ideal detectors for triggering muons, able to unambiguously identify the relevant bunch crossing a muon track is associated to. Each component is discussed in more detail in [10, 116].

Figure 25 shows the expected momentum resolution of the muons based on tracker information only, muon chamber information only and a combination of tracker and muon chamber information. At low muon p_T the resolution is driven from the reconstruction of muons in the tracker, however as the p_T of the muon increases then the combination of information from the tracker and muon chambers results in improved resolution.

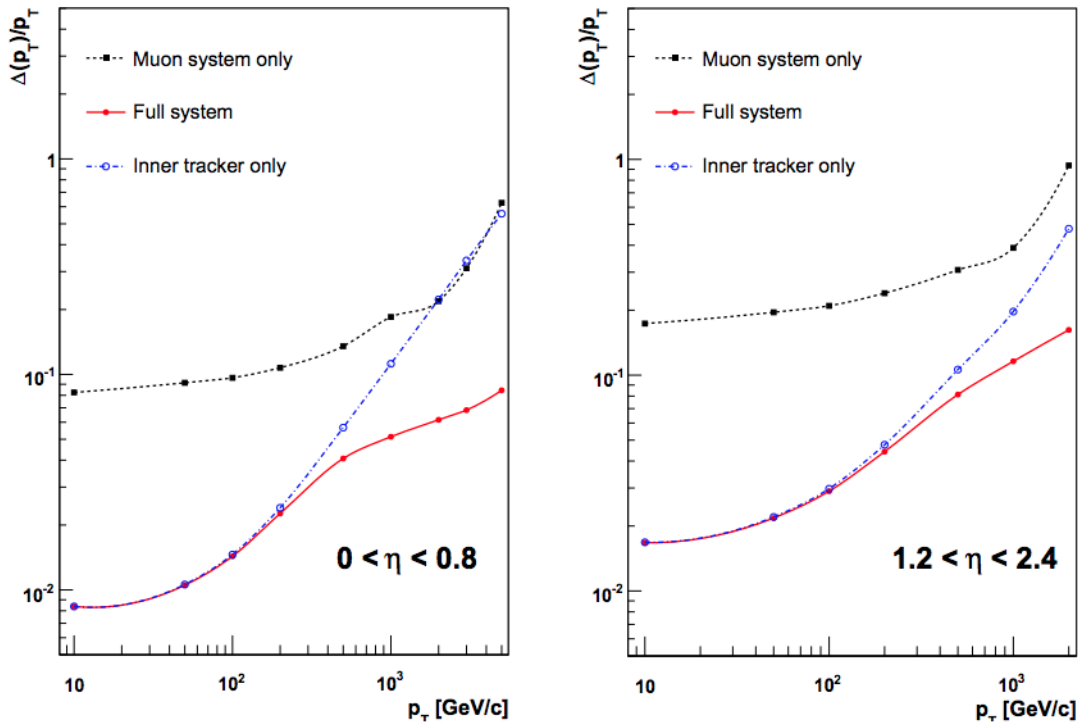


Figure 25: The muon p_T resolution as a function of p_T using the tracking only, muon subdetectors only or both. Figure taken from [10].

3.3.6 Reducing the volume of data using triggering

The CMS detector produces events at an approximate rate of 40 MHz. If each reconstructed event has a size of ~ 1 MB, then data would be produced at a rate of 40 TB s^{-1} . It is not feasible to fully reconstruct and store the full rate of data production so a method of data reduction, called triggering, is used. Triggering decides on an event-by-event basis whether a collision is likely to be highly-energetic and interesting as opposed to an elastic collision or low-energy QCD process. There are two types of triggering systems at CMS, the Level-1 hardware triggers (L1T) and the High-Level software triggers (HLT).

The L1T must decide within $3.8 \mu\text{s}$ of an event occurring whether it should be rejected or tentatively accepted. The schematic workflow of the L1T is shown in Fig. 26. It consists of sophisticated algorithms, inserted into large field-programmable gate arrays (FPGAs), the inputs and outputs of which are connected by high-speed optical links. The fast, programmable

nature of the L1T hardware, means that further improvements can be brought to the performance of the trigger [117]. The L1T uses only basic information from the detector, known as trigger primitives, to make a decision. These trigger primitives are formed from the calorimeter deposits and the hits in the muon detectors. There are two trigger subsystems to the L1T, the calorimeter trigger and the muon trigger. These access the trigger primitives from different subdetectors and operate independently, with the exception of information passed from the calorimeter trigger to the muon trigger for the computation of muon isolation. The outcome of both trigger components is then passed to the global trigger which makes the final decision on the event. Events are stored in a buffer while for L1T processing proceeds and if the decision is not returned in time the event is deleted. To this end, the travel time between detector and computer farm needs to be as short as possible, and so the processing facilities are located in the underground service cavern next door to the detector. The L1T reduces the event rate from 40 MHz to 100 kHz, which provides enough latency for the events to be fully reconstructed before being processed by the HLT. Approximately 200 L1T algorithms form the seeds for the HLT.

Events passing the L1T are passed to the HLT located on a computer cluster at the surface. The software-based HLT, with access to the L1T seeds and full event reconstruction, using algorithms similar to those presented in Ch. 5, reduces the final physics production rate to 1 kHz. It can classify the events into data sets with similar topologies, for example into sets containing a single isolated electron or sets that have high jet multiplicities. Due to the nature of software, the HLT algorithms can be easily updated with new triggers of more efficient algorithms.

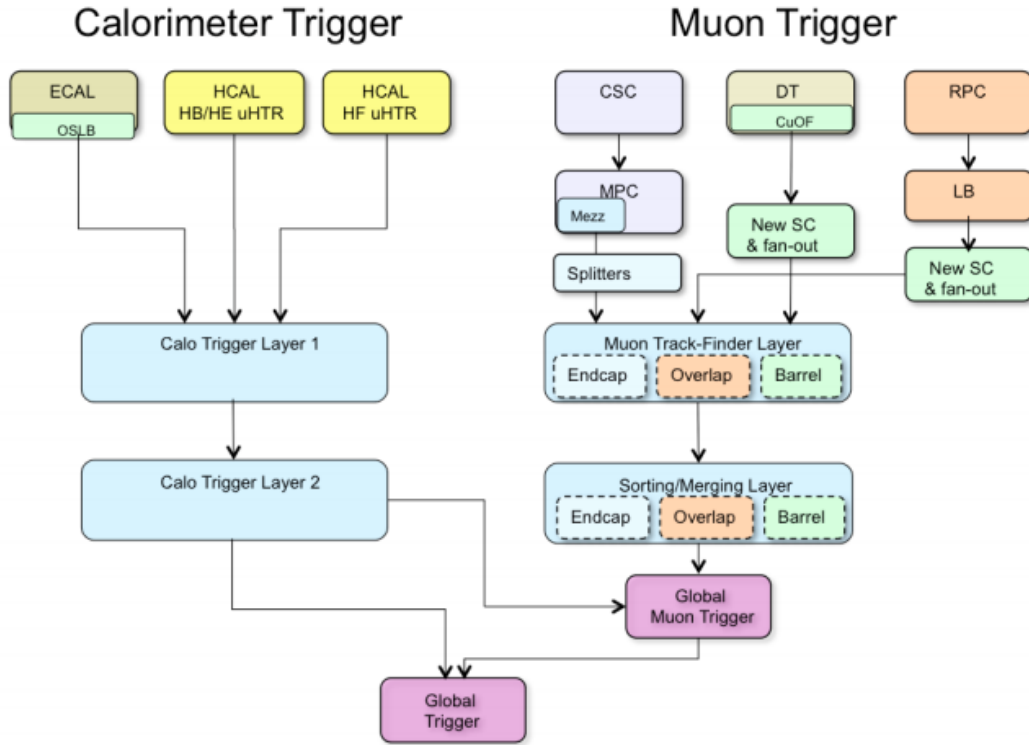


Figure 26: A schematic representing the process flow of the L1T. On the left is the calorimeter trigger and on the right the muon trigger which operate on independent trigger primitives. The final decision is taken by the global trigger using all information available [118].

3.3.7 Computational power

The field of particle physics is unique in the way in which it shares data sets. By considering just the samples stated in Tab. 4, there are, conservatively, one billion events leading to a total simulated data set size of > 1 PB in just this thesis alone. As well as the simulated samples, the data sets produced by the LHC must also be stored and together this is far too much for a single storage site to be viable or for every analysis to use a set of privately generated simulations and so the data sets must be shared.

Not only must data storage be shared across the collaboration, but also the processing power. By taking the same number of events, it would take approximately two millenia to simulate the events (at 1 event per minute) and approximately eight months to run the analysis code for this thesis (at 20 events per second) on a single processing core. Waiting almost two thousand years for data samples to be generated is clearly not feasible, therefore the data simulation and processing needs tens of thousands of cores.

The Worldwide LHC Computing Grid (WLCG) was created to share storage and processor resources. The WLCG is a tiered system with all raw data being stored at a central Tier-0 computing facility and subsequently distributed to Tier-1 and Tier-2 sites in the countries and institutes affiliated to CERN. The complete set of raw data is stored at the CERN Tier-0 data centre, where the combined permanent archive stored on tape breached 200 PB on the 29th June 2017. In terms of online storage the data centre at CERN holds 45 PB with a further 5.5 PB extension at the Wigner Research Centre for Physics in Budapest. The 100 000 processor cores at the data centre deal with 1 PB of data every day.

The large Tier-1 facilities store a large fraction of the data and simulation sets so that there are multiple redundancies for every data set. The 13 Tier-1 facilities are the main workhorses of the WLCG performing large-scale reprocessing of the raw data sets and the generation of simulated events.

Tier-2 sites, of which there are 155, store data sets distributed to them from Tier-1 facilities and specific data sets that are useful for locally based analysis teams. They provide a moderate number of processor cores for computational tasks, most generally for the production of simulated events. The distributed nature of the file system allows for any person to access data sets from any site which has it stored and to use free computational power anywhere that is part of the WLCG, increasing the overall computational efficiency.

3.3.8 Data collection in 2016

As already stated in Sec. 3.2, an integrated luminosity of 35.9 fb^{-1} was collected by CMS in 2016. To further classify the data used in this thesis, Tab. 3 shows the data collected with respect to the run period. A run period refers to periods of similar data taking throughout the year. Run period A is used for the commissioning of the detector. Run periods often change after a short technical stop has occurred. The initial data sets used in this thesis are classified as containing either a single electron or a single muon. The HLT trigger selections reduce the total number of events used from the initial data sets significantly, decreasing computational time. In the $e+\text{jets}$ channel the trigger requires there to be an electron with a $p_T > 32 \text{ GeV}$, a $|\eta| < 2.1$ and to pass a set of tight identification requirements similar to those seen in Sec. 5.2.1. In the $\mu+\text{jets}$ channel an isolated global muon or isolated tracker muon is required with a $p_T > 24 \text{ GeV}$ and a $|\eta| < 2.4$.

Table 3: The sets of single electron and single muon data collected during 2016 split into individual runs. A total luminosity of 35.9 fb^{-1} of data was collected.

Channel	Data set	Run ranges	Luminosity (fb^{-1})
SingleElectron(Muon)	Run2016B-03Feb2017_ver2-v2	272007–275376	5.79
	Run2016C-03Feb2017-v1	275657–276283	2.57
	Run2016D-03Feb2017-v1	276315–276811	4.25
	Run2016E-03Feb2017-v1	276831–277420	4.01
	Run2016F-03Feb2017-v1	277772–278808	3.10
	Run2016G-03Feb2017-v1	278820–280385	7.54
	Run2016G-03Feb2017-ver2(3)-v1	280919–284044	8.61
Total integrated luminosity			35.87

CHAPTER 4

Modelling proton-proton interactions

For any measurement in particle physics, sets of simulated samples need to be generated in order to compare the experimental data to theoretical models. These models include well known SM decay processes such as $t\bar{t}$ production, much rarer processes predicted by the SM such as $t\bar{t}$ production in association with a H boson, and in the case of a search for new physics, a plethora of processes predicted by beyond the SM theories.

These different processes are modelled by event generators, following a sequence of discrete steps. First of all, a possible hard scattering interaction is generated for a process at LO, or if possible NLO. If short-lived resonances are formed, their associated decays are also viewed as part of the hard process. Additional parton interactions can be included at LO and soft radiation added to the initial and final states, in a process known as the parton shower. The parton shower is then hadronised into colourless states and decayed, with the proton remnants and additional parton interactions forming the underlying event. Figure 27 shows diagrammatically an example simulated collision, where the hard scattering process is shown in red, the soft radiative processes in blue, the additional parton interactions in purple, the fragmentation in light green and hadronisation and decay in dark green. Photon radiation is shown in yellow. Finally, the interactions of the particles produced with the detector can be modelled and the detector response applied.

4.1 Hard interaction

The hard scattering process describes the interaction of two partons in the collider. It is modelled through Eq 2.31, with the PDFs taken from data convoluted with the partonic cross section. The partonic cross section is calculated from the matrix-element of the interaction, directly taken from the Feynman diagrams through the Feynman rules. If a short-lived resonance is formed as a result of the interaction, then its decay is also considered as part of the hard interaction. Many matrix-elements can already be calculated to NNLO or higher, however the generation of the hard scattering processes can currently only be calculated to LO or NLO. This is due to the combinatorics when taking into account all possible Feynman diagrams.

There are a variety of matrix-element generators available, two of the most popular in high energy particle physics being MadGraph5_aMC@NLO matrix-element generator (MG5_aMC@NLO) [35] and PPositive Weight Hardest Emission Generator matrix-element

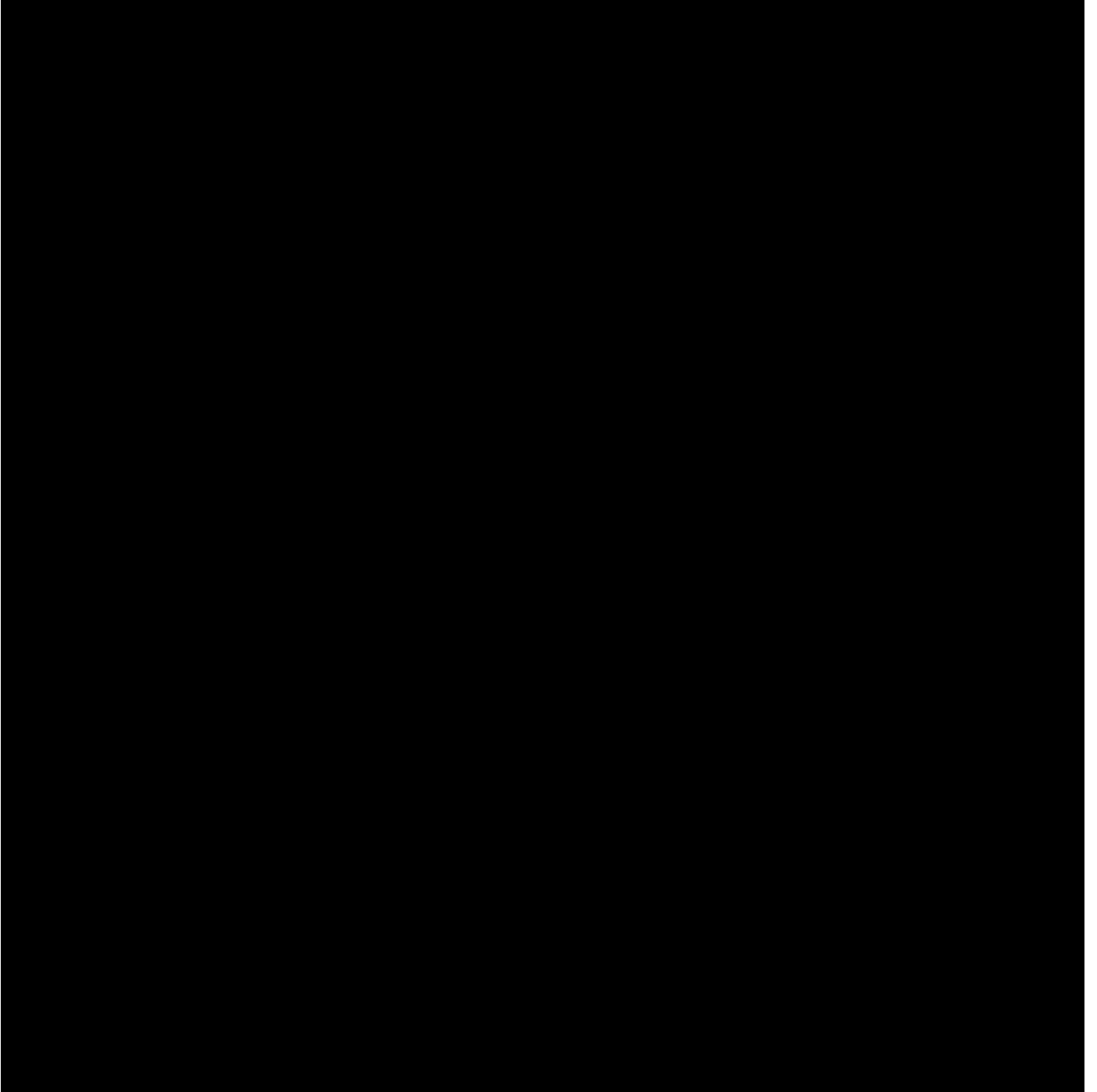


Figure 27: An example simulation of a proton-proton collision showing all the various processes that must be modelled. Starting with the hard interaction in red calculated using perturbation theory, then dressing with additional soft processes in blue forming the parton shower. Additional parton interactions are added, shown in purple, each with a respective parton shower. Finally the parton shower is fragmented and hadronised into detectable, colourless hadrons, which can in turn decay and radiate. This fragmentation and hadronisation processes are shown in light green and green respectively. Photon radiation is shown in yellow. Figure taken from [119].

generator (POWHEG) [120–122]. Both the matrix-element generators can model a process with additional partons. In MG5_aMC@NLO, the number of additional partons in the process depends on the complexity of the process, e.g. for NLO $t\bar{t}$ production, an additional two partons can be generated. It does this by adding real diagrams based on LO or NLO contributions from lower order processes. In POWHEG, virtual processes are included in the matrix element for additional partons, up to one loop. This means that for $t\bar{t}$ production, one less additional parton can be produced than for MG5_aMC@NLO.

4.2 Additional interactions and soft processes

The most common hard process at the LHC is elastic gluon-gluon scattering, the cross section of which, diverges as the gluons become soft and collinear. This means that below some momentum transfer threshold (~ 5 GeV) the inclusive jet production cross section from elastic gluon-gluon scattering is larger than the total inclusive proton-proton cross section [123, 124], which infers that for every collision, more than one interaction is occurring. These are known as multiple parton interactions (MPI). The number of additional interactions present is proportional to the impact parameter of the colliding protons. A pair of protons colliding head on contains more additional interactions than a pair that collide obliquely. Once the hard process has been simulated from the matrix-element calculations, the possible MPIs from the other initial state partons are added. All the hard interactions present in the simulation must then be dressed with soft emissions.

The partons produced in the hard interaction and MPIs may radiate energy in the form of additional partons. The additional partons come in the form of initial and final state radiation (ISR and FSR) and are included in the simulation either as part of the hard scattering process or as part of the parton shower. Figure 28 shows the four possible ways in which a parton can be radiated. Firstly, a quark can radiate a gluon, secondly a gluon can pair-produce a $q\bar{q}$ pair and finally there are two colour configurations of gluon splitting.

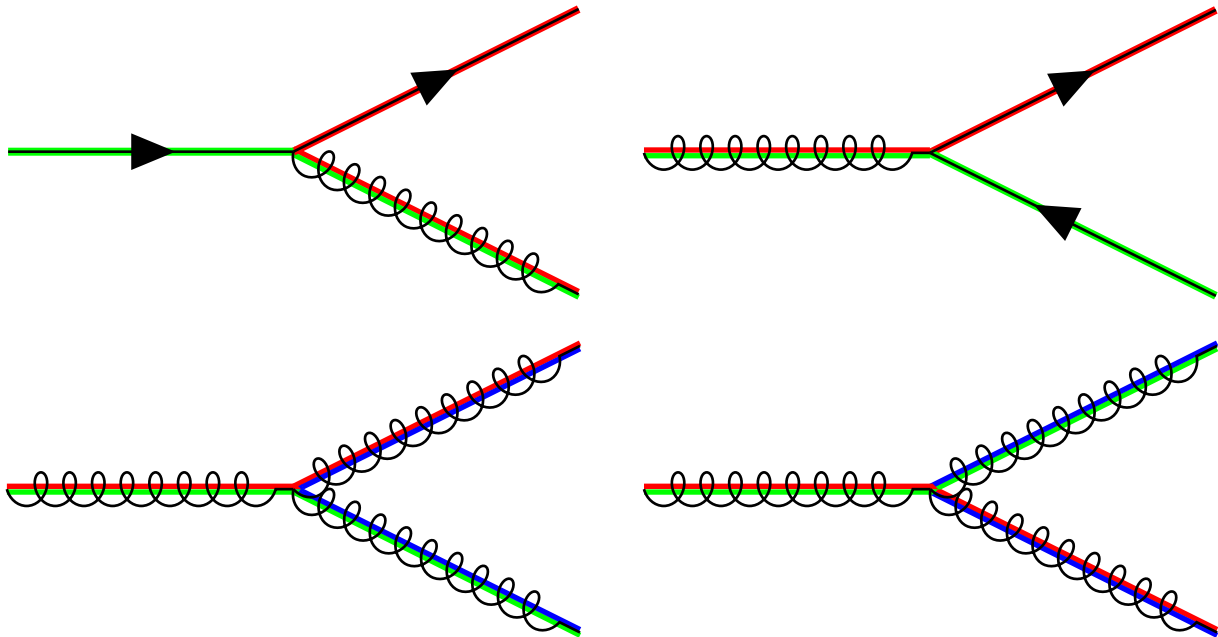


Figure 28: The four possible processes to strongly split a parton in the parton shower. The top left panel shows gluon emission, the top right shows $q\bar{q}$ pair production and the bottom two panels the colour configurations of gluon splitting. In addition, photon brehmsstrahlung from the quarks can occur but is not considered here.

For FSR, the partons produced from the matrix-element calculation are split recursively until the energy of each remaining parton reduces to approximately 1 GeV, whereupon hadronisation effects become non-negligible.

In contrast, the ISR is modelled differently. The longitudinal momentum fractions, x_1 and x_2 , of the incoming partons are taken from the PDF at the energy scale of the hard interaction and evolved backwards in time. It is important to avoid using a possible non-perturbative parton initiating the hard scatter, therefore the factorisation scale is used as a cutoff. Partons that carry a momentum less than the factorisation scale are absorbed into the PDF, and those that carry a momentum above the scale are allowed to participate in the interaction. This ensures no divergences in the matrix-element due to collinear or soft gluons. For every ISR emission, the parent parton must have the combined energy of the daughter parton and radiation. As there are fewer available partons carrying an increased fraction of the proton momentum, as seen in the PDF distribution, a suppression weight must be applied. The suppression weight is derived from the Sudakov form factors [125]. In practice, the application of ISR and MPI occur completely and in tandem, before the FSR is modelled.

The collective spray of partons produced by the hard process, MPI, ISR and FSR is called the parton shower. The evolution of the splittings in the parton shower is defined by ordering variables. One such shower ordering variable is the p_T^2 of the shower propagators, where the MPIs, ISR and FSR are interleaved with one another into one common decreasing p_T^2 sequence [126]. The p_T^2 ordering is illustrated in the left panel of Fig. 29, where the hard interaction occurs with $p_{T,\max}^2(A/B)$ and the subsequent evolution follows $p_{T,\max}^2 > p_{T,1}^2 > p_{T,2}^2 > p_{T,3/4}^2$.

Another way to evolve the shower is by using an angular ordering [127]. This angular ordering requires any future splitting to occur within a cone defined by the angular radius between the daughter partons from the previous splitting, also illustrated in the right panel of Fig. 29. These showers often produce wide-angle soft splittings before the main hard splitting. Two widely used models for the parton shower are PYTHIA [128] and HERWIG++ [129], where PYTHIA uses a p_T -ordered shower and HERWIG++ an angular-ordered shower.

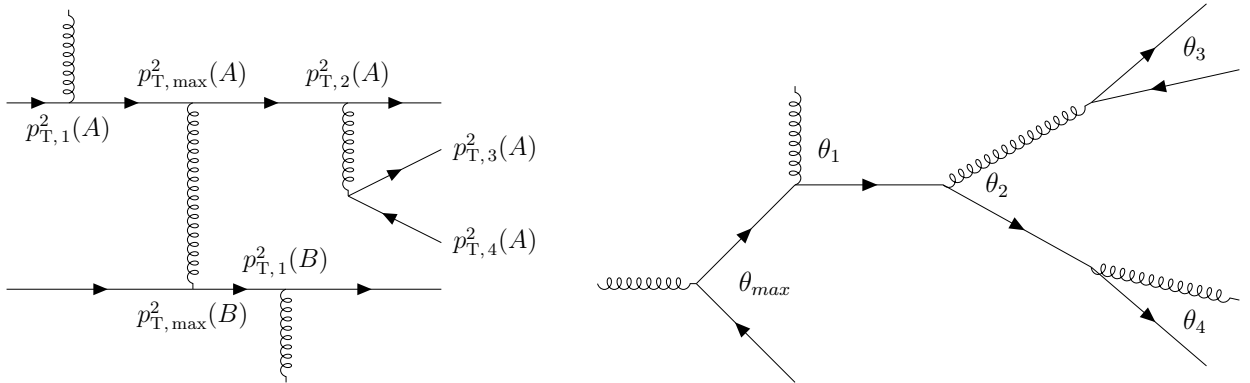


Figure 29: The left panel shows an example p_T^2 ordered parton shower. For each participating hadron (A/B) the shower is centered on the hard interaction and decreases in hardness the further from this point, such that $p_{T,\max}^2(A) > p_{T,1}^2(A) > p_{T,2}^2(A) > p_{T,3/4}^2(A)$ and $p_{T,\max}^2(B) > p_{T,1}^2(B)$. The right panel shows shower ordering by angle such that the opening angle decreases for every consecutive splitting $\theta_{\max} > \theta_1 > \theta_2 > \theta_{3/4}$.

4.3 Matching matrix-elements to parton showers

Parton showers and higher-order matrix-element calculations describe the same physics and so need to be combined consistently. The soft emissions added from the parton shower should

interface with the matrix-element based radiation with no gaps or overlaps for a specific, fixed final state parton multiplicity. This problem is solved by a process known as matching.

One method of matching, known as the MLM-method [130], after its author M. L. Mangano, matches the final state partons produced by the matrix-element calculation to parton jets created by clustering the showered partons. The angular distance from each final state parton to each parton jet, starting from the parton with the leading p_T , is calculated and is said to be matched if it is less than a threshold value. If the matching is successful the parton jet is removed from further iterations. All final state partons must be matched to parton jets, otherwise the event is rejected. When simulating the hard interaction process, including contributions with up to N additional partons, events in samples simulated with n additional partons, where $n < N$, are suppressed if they do not contain exactly n matched parton jets. This is because these events are present and more accurately modelled in the $n + 1$ production sample and thus double counting is avoided. All the production samples generated using discrete additional parton multiplicities are subsequently merged.

When matching and merging at NLO the additional parton jet from the matrix-element calculation is the hardest emission and thus defines the upper shower scale level. While it is easy to interface this with showers ordered to some p_T , care must be taken with respect to an angular-ordered shower. Angular-ordered shower emissions with a p_T greater than the additional matrix-element parton jet need to be suppressed. The FxFx-method [131], extends the MLM-method of merging to matrix-elements calculated to NLO. The MLM-method is used when matching LO MG5_aMC@NLO to PYTHIA and the FxFx-method is used when matching NLO MG5_aMC@NLO to PYTHIA.

When POWHEG models hard scattering processes at a large energy scale, the number of high- p_T soft emissions is overestimated [132, 133]. It is resolved by introducing a p_T dependent damping factor,

$$D = \frac{h_{\text{damp}}^2}{p_T^2 + h_{\text{damp}}^2}, \quad (4.1)$$

where h_{damp} is the damping parameter, usually set to the hard scattering energy scale. D effectively sets the upper- p_T limit for the first additional emission and hence the upper shower scale level.

4.4 Fragmentation and colour reconnection

The fragmentation process models how final state partons from the parton shower hadronise into colourless states. There are two general methods in use for the fragmentation, the Lund string model [128, 134] and the cluster model [129].

4.4.1 Lund string model

Due to the self interactions of gluons, the strong field between two quarks can be represented by a string. As the quarks move apart the potential energy in the string increases linearly by 1 GeV fm^{-1} . At some point the colour field contains enough potential energy that it is favourable to pair-produce light quarks, which combine with the initial quarks to form two new colourless particles, breaking the string. The string is stretched between quark endpoints by a number of gluons. Each gluon results in a kink to the string that depends on the kinematic properties of the gluon. The left panel of Fig. 30 shows the kink in a string produced by the addition of a gluon and the right panel shows the fragmentation of the string. As time evolves, if the $q\bar{q}$ pairs do not have enough potential in the colour field to pair-produce they end up bound to each other forming the colourless mesons of the final state. It is also possible for the

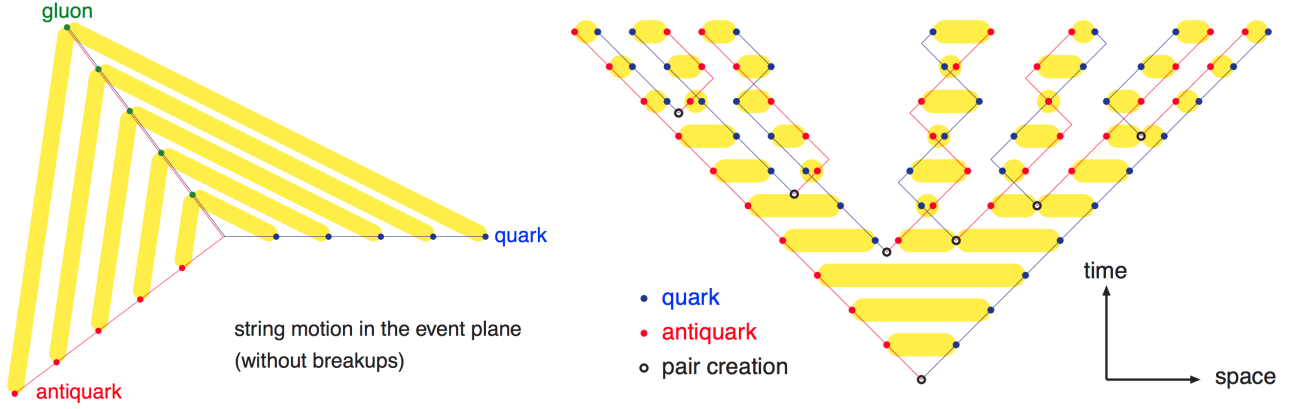


Figure 30: The left panel shows the string between the quark-antiquark colour dipole stretched by a gluon and the right panel shows for each string how fragmentation occurs with time. Lorentz boosts in the hadrons are shown by oscillation frequency. Figures taken from [135].

potential in the colour field to be great enough to pair-produce diquark-diantiquarks instead of $q\bar{q}$ with the end result being the formation of baryons instead of mesons. This approach is used in the PYTHIA parton shower simulation.

The string dynamics currently assume that the colour-flow modelled in the evolution of the parton shower from the hard interaction is independent of the colour-flow of the other parton showers produced by the MPIs. Indeed, each individual MPI is assumed to operate in independent colour spaces. In reality, all the individual interactions are superimposed on each other, with soft gluons connecting all the final state partons from all sources, redistributing colour across the full parton shower. In the Lund string model this can have the effect of reorganising the strings produced in the parton shower. This is modelled in simulation by a process known as colour reconnection (CR), which reconfigures the colour strings after the parton shower, but before the fragmentation. Three different models which can be implemented using PYTHIA are described here.

MPI-based model

The MPI-based CR model [136] calculates the reconnection probability for all gluons from lower- p_T interactions to be inserted into the colour strings of a higher- p_T interaction. Once all reconnections between MPIs have been identified, the configuration in which the total colour string length is minimised is selected. The PYTHIA parton shower simulation uses the MPI-based CR scheme as default, where it is assumed that the lifetime of the top quark is long enough to shield the top decay products from the effects of CR. A second MPI-based model exists where the top quark decays before the CR is applied and as such the top decay products are able to reconnect. This is known as MPI-based CR with early resonance decays (erdON).

QCD-inspired model

An alternative CR model is the QCD-inspired model [137]. In this model, instead of the $SU(3)$ colour indices (r, g, b) being applied to individual partons, a set of nine indices, which describe the possible colour states of 2-parton and 3-parton combinations, are used. A corresponding set of anticolour indices is used for antiquarks and a gluon contains one of each index. Reconnections are possible when two partons have a matching colour and anticolour index. The set of possible reconnections encompasses every possible reconnection from the default MPI-based CR model as well as reconnections possible between partons with accidentally matching colour indices. As with the MPI-based CR model, a reconnection is performed if it minimises the total

QCD string length.

Gluon move/flip model

The gluon move/flip model [136], moves all final state gluons between colour strings so that the total string length is minimised. However, by moving only gluons, no quarks are able to reconnect and so a subsequent procedure is applied in which two colour segments are flipped between colour strings. As with moving the gluons, the solution which minimises the total string length is used.

4.4.2 Cluster fragmentation model

In the cluster model of fragmentation implemented in the HERWIG++ parton shower simulation, the parton shower is evolved until the partons reach an energy scale of $\Lambda_{\text{QCD}} \approx 4 \text{ GeV}$, where hadronisation effects become non-negligible. The final state gluons must have an energy at least twice that of the invariant mass of the lightest quark and are forced to split into a $q\bar{q}$ pair. The colour singlet states formed by the resulting colour connected partons are formed into clusters, as shown in Fig. 31. The cluster model assumes the principle of colour-preconfinement which states that the distribution of the mass of the clusters is independent of the centre-of-mass energy and the hard scattering process involved. Most clusters are distributed at low masses and can be regarded as excited hadron resonances and decayed into observed hadrons, however some clusters are too massive for this to be realistic and split in a process called cluster fission. If fission is required, a $q\bar{q}$ pair of u, d or s flavour, is taken from the vacuum and the cluster split such that one of the original quarks is contained in each new cluster. Fissions continue until all clusters are accepted by the mass parameters of the model. The final process of the fragmentation involves extracting a $q\bar{q}$ or $qq\bar{q}\bar{q}$ from vacuum in each cluster, forming a pair of hadrons, which in turn may decay into lighter hadrons.

4.5 Detector modelling

So far, only the modelling of the physics process itself has been discussed, before any interaction with the detector has been performed. Indeed, it is very useful to compare measurements at this level, the particle level, as it allows direct comparisons between theoretical models and experimental measurements which have had the detector response removed. The removal of the detector response is discussed in Ch. 7. This means measurements performed to particle level with different detectors such as CMS and ATLAS can be directly compared with each other and to theoretical predictions.

In order to provide the mapping between experiment and particle level, the detector response needs to be modelled accurately. To do this a simulation toolkit called GEometry ANd Tracking (GEANT4) [139] is used. The GEANT4 package builds the physical layout of the experiment, including all detectors, absorbers, electronics and structural components surrounding the central interaction point. The magnetic field strength at all points in the detector is modelled. The particles are tracked, in iterative time steps, from the collision point outwards through the active and non-active material, modelling respective interactions and decays. In the active material, the response of all individual components is modelled in a process known as digitisation. The digitisation takes into account any inefficiencies in the active material and returns an output consistent with that seen from measured data. A full description of the CMS detector is given in Ch. 3.

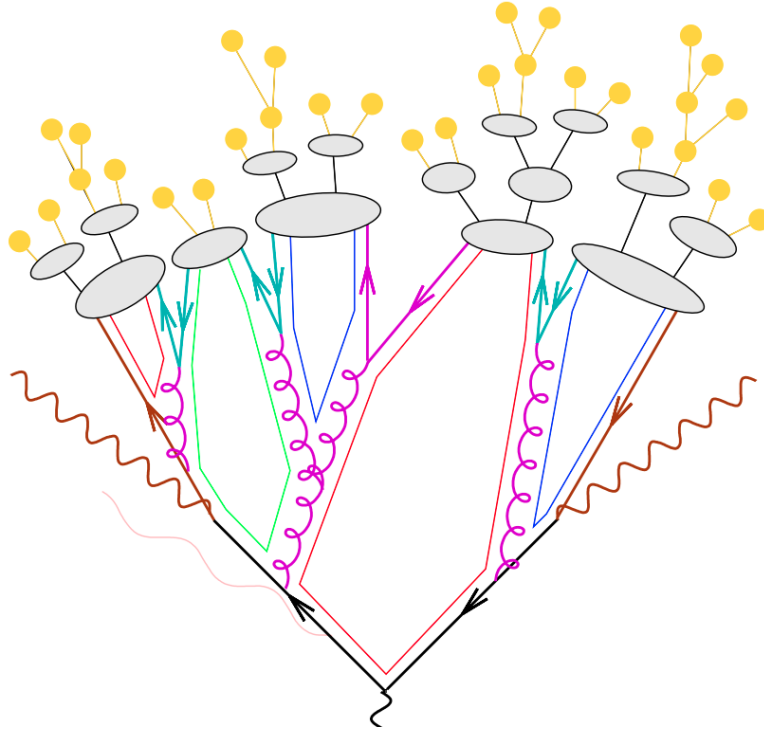


Figure 31: A sketch of the cluster model of fragmentation and hadronisation. Colour singlet clusters are generated from the final state quarks and gluons forced to split into a $q\bar{q}$ pair. These colour singlet clusters can be unnaturally massive and fission occurs until the clusters represent excited hadron resonances which then hadronise. Figure adapted from [138].

4.6 Production of top quark pair and background models

Table 4 shows the complete set of simulations used for the differential cross section analysis. There are four different $t\bar{t}$ production models studied. Two $t\bar{t}$ samples are simulated with the POWHEG (v2) matrix element generator [140], where one sample is interfaced with PYTHIA (v8.212), using the CUETP8M2T4 tune [82], and the second with the HERWIG++ (v2.7.1) using the EE5C tune [141]. These are labelled POWHEG+PYTHIA and POWHEG+HERWIG++ respectively. Two additional simulated $t\bar{t}$ samples are produced with the MG5_aMC@NLO (v2.2.2) generator. One is used to generate events at LO accuracy with up to three additional partons and the second to NLO accuracy with up to two additional partons. Both are also interfaced with PYTHIA (v8.212) using the CUETP8M1 tune [142] and are labelled MG5_aMC@NLO-LO and MG5_aMC@NLO-NLO respectively. The MLM jet-parton matching algorithm is used to match the matrix-element to the parton shower in the LO simulation and the FxFx algorithm in the NLO simulation. The POWHEG+PYTHIA model is taken as the central model used in this thesis.

All $t\bar{t}$ production samples are simulated with the top quark mass set to 172.5 GeV. For NLO samples the NNPDF30_nlo_as_0118 PDF set [143] is used, while for LO samples the NNPDF30_lo_as_0130 set [143] is used. The samples are normalised to a NNLO $t\bar{t}$ cross section of 832 pb, as seen in Sec. 2.7.1. It is important to compare multiple $t\bar{t}$ generators in order to find the current most suitable description of top quark production and decay, and to identify any discrepancies in the models.

Table 4 also shows the dominant background samples from single top quark production, vector boson production with associated jets and multi-jet QCD production. All the single top quark processes produced via the t -channel and in association with a W boson are generated with POWHEG [144, 145] interfaced with PYTHIA and are normalised to cross sections that are

calculated to NLO precision [146, 147]. Samples of W and Z boson production in association with jets, with leptonic final states, are generated at LO using MG5_aMC@NLO. Separate samples are generated with exactly one, two, three, and four additional jets to ensure a large set of background events. These samples are normalized to their NNLO cross sections [148]. Multijet QCD events are generated using PYTHIA for both the matrix-element calculations and parton shower/hadronisation. Only multijet QCD events with large electromagnetic activity or that contain a muon are generated in order to maximise the available sample size.

The simulated samples need to contain enough events to ensure they are not affected by statistical fluctuations when comparing to the data. The equivalent luminosity produced for each simulation sample is shown in the final column of Tab. 4 and should ideally be comparable to or larger than the luminosity of the data collected. The total luminosity collected in 2016 is 35.9 fb^{-1} , so most of the signal and background samples contain enough events. The simulated multijet QCD samples are not particularly well modelled and can also suffer from a lack of generated events and so a data-driven multijet QCD estimate is employed, described in Sec. 6.5. The multijet QCD samples are still required in the creation of the estimate however.

Samples are also generated for variations of some of the generator parameters in POWHEG+PYTHIA. These are shown in Tab. 5 and are used as part of the estimation of the theoretical uncertainties due to the modelling. Other theoretical uncertainties in the modelling are calculated as event weights and stored in the central POWHEG+PYTHIA sample. The CMS detector response for all simulated samples is modelled using GEANT4.

Table 4: The set of simulated samples used in this thesis. They include four different signal $t\bar{t}$ production samples and samples for the background production. The backgrounds are single top quark production, vector boson production in association with jets and multijet QCD. The cross sections for the multijet QCD are shown in the form $X * Y$, where X refers to the cross section and Y refers to the filter efficiency for the specific type of multijet QCD being selected (e.g. muon enriched).

	Matrix-element	Parton shower	Underlying event tune	Cross section (pb)	Events ($\times 10^6$)	Equivalent L (fb $^{-1}$)
Models of $t\bar{t}$ production						
POWHEG+PYTHIA	POWHEG	PYTHIA	CUETP8M2T4	831.76	154.9	186.3
POWHEG+HERWIG++	POWHEG	HERWIG++	EE5C	831.76	59.2	77.1
MG5_aMC@NLO-LO	MG5_aMC@NLO	PYTHIA	CUETP8M1	831.76	10.1	12.2
MG5_aMC@NLO-NLO	MG5_aMC@NLO	PYTHIA	CUETP8M1	831.76	30.4	36.6
Models of single top background production						
Single top t -channel	POWHEG	PYTHIA	CUETP8M1	136.02	66.9	492.0
Single anti-top t -channel	POWHEG	PYTHIA	CUETP8M1	80.95	38.8	479.4
Single top tW -channel	POWHEG (v1)	PYTHIA	CUETP8M1	35.6	7.9	223.2
Single anti-top tW -channel	POWHEG (v1)	PYTHIA	CUETP8M1	35.6	7.9	222.8
Single top/anti-top s -channel	MG5_aMC@NLO	PYTHIA	CUETP8M1	6.35	0.6	98.1
Models of vector boson background production						
Drell-Yan + 1 jet	MG5_aMC@NLO	PYTHIA	CUETP8M1	1016	62.6	61.6
Drell-Yan + 2 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	331.4	20.0	60.3
Drell-Yan + 3 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	96.36	5.9	60.8
Drell-Yan + 4 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	51.4	4.2	81.7
W boson + 1 jet	MG5_aMC@NLO	PYTHIA	CUETP8M1	9493	45.4	4.8
W boson + 2 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	3120	60.2	19.3
W boson + 3 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	942.3	59.1	62.7
W boson + 4 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	524.2	30.0	57.2
Models of multijet QCD background production						
Muon enriched QCD (20-30)	PYTHIA	PYTHIA	CUETP8M1	558528000 * 0.0053	30.6	0.01
Muon enriched QCD (30-50)	PYTHIA	PYTHIA	CUETP8M1	139803000 * 0.01182	30.0	0.02
Muon enriched QCD (50-80)	PYTHIA	PYTHIA	CUETP8M1	19222500 * 0.02276	19.8	0.04
Muon enriched QCD (80-120)	PYTHIA	PYTHIA	CUETP8M1	2758420 * 0.03844	23.6	0.2
Muon enriched QCD (120-170)	PYTHIA	PYTHIA	CUETP8M1	469797 * 0.05362	8.0	0.3
Muon enriched QCD (170-300)	PYTHIA	PYTHIA	CUETP8M1	117989 * 0.07335	17.4	2.0
Muon enriched QCD (300-470)	PYTHIA	PYTHIA	CUETP8M1	7820.25 * 0.10196	49.0	61.4
Muon enriched QCD (470-600)	PYTHIA	PYTHIA	CUETP8M1	645.528 * 0.12242	19.0	240.1
Muon enriched QCD (600-800)	PYTHIA	PYTHIA	CUETP8M1	187.109 * 0.13412	10.0	397.7
Muon enriched QCD (800-1000)	PYTHIA	PYTHIA	CUETP8M1	32.3486 * 0.14552	19.8	4199.3
Muon enriched QCD (1000-Inf)	PYTHIA	PYTHIA	CUETP8M1	10.4305 * 0.15544	13.4	8264.9
Electron enriched QCD (20-30)	PYTHIA	PYTHIA	CUETP8M1	557600000 * 0.0096	9.2	0.002
Electron enriched QCD (30-50)	PYTHIA	PYTHIA	CUETP8M1	136000000 * 0.073	6.8	0.0007
Electron enriched QCD (50-80)	PYTHIA	PYTHIA	CUETP8M1	198000000 * 0.146	45.2	0.02
Electron enriched QCD (80-120)	PYTHIA	PYTHIA	CUETP8M1	2800000 * 0.125	76.5	0.2
Electron enriched QCD (120-170)	PYTHIA	PYTHIA	CUETP8M1	477000 * 0.132	77.8	1.2
Electron enriched QCD (170-300)	PYTHIA	PYTHIA	CUETP8M1	114000 * 0.165	11.5	0.6
Electron enriched QCD (300-Inf)	PYTHIA	PYTHIA	CUETP8M1	9000 * 0.15	7.4	54.6
bc to E QCD (30-80)	PYTHIA	PYTHIA	CUETP8M1	159068000 * 0.00255	15.3	0.04
bc to E QCD (80-170)	PYTHIA	PYTHIA	CUETP8M1	3221000 * 0.01183	14.9	0.4
bc to E QCD (170-250)	PYTHIA	PYTHIA	CUETP8M1	105771 * 0.02492	9.7	3.9
bc to E QCD (250-Inf)	PYTHIA	PYTHIA	CUETP8M1	21094.1 * 0.03375	9.8	13.7

Table 5: The set of simulated samples associated with the uncertainties in the POWHEG+PYTHIA modelling. Other uncertainties are modelled by the use of event weights in the central POWHEG+PYTHIA simulation.

Variation	Events ($\times 10^6$)	Equivalent L (fb^{-1})
ISR down	59.0	70.9
ISR up	59.0	71.0
FSR down	59.3	71.3
FSR up	59.2	71.2
Underlying event down	58.3	70.1
Underlying event up	59.0	70.9
Matching scale (h_{damp}) down	58.2	69.9
Matching scale (h_{damp}) up	58.9	70.8
CR MPI-based erdON	59.9	72.0
CR QCD-inspired	59.6	71.7
CR Gluon move	59.0	71.0
Top quark mass (169.5 GeV)	59.5	70.4
Top quark mass (175.5 GeV)	59.4	71.4

CHAPTER 5

Reconstructing the physics objects

Whether the components of an event are measured from a detector or simulated from a model, the charge and energy deposits need to be reconstructed into physics objects. In the CMS experiment, the reconstruction of objects from the calorimeter energy deposits and charge deposits in the tracker and muon chambers is done by an algorithm called particle flow (PF) [149]. These collections of deposits are reconstructed into likely electrons, photons, muons, charged and neutral hadronic jets and missing p_T and are known as PF candidates. The PF candidates undergo further, stricter identity requirements depending on the needs of the analysis, into high-purity collections of object candidates.

5.1 The particle flow algorithm

The PF algorithm is used to reconstruct the raw deposits from the detector into PF candidates. The first step is to form the basic building blocks of the algorithm, known as PF elements. The PF elements are the charged particle tracks and vertices along with the calorimeter clusters and muon system hits. Each PF candidate is reconstructed using one or more of these elements. The combination of these correlated elements, allows for the reconstruction and classification of the final state particles present in an event.

5.1.1 Particle flow elements

Tracks are built in an iterative process. Firstly, a track seed is generated for a triplet of hits in the pixel detector, with a few hits compatible with a charged-particle trajectory. Next, additional hits are added, layer-by-layer, to the current track candidate based on the track seed trajectory, extrapolated via a Kalman filter [150]. New track candidates are formed in this process when multiple hits match the extrapolation and track candidates with missing hits are also kept. Once the outer tracker layer has been reached, possible duplicates are avoided by requiring each pair of track candidates to be composed of at least 50% of unique hits. If not, then the candidate with the most missing hits is dropped and if they are the same length then the one with the highest χ^2 is dropped. The final track fitting is then performed, using all associated track hits to determine the charged-particle properties (vertex, p_T , etc...). Additional track quality requirements are imposed which include a minimum p_T of 0.9 GeV, the

number of missing hits, the goodness-of-fit χ^2 of the track fit and the distances to the beamline and a reconstructed primary vertex. Once a track candidate is accepted its component hits are removed from the collection of tracker hits, and the process repeats, until no more track candidates can be found.

Primary vertices are reconstructed by clustering the reconstructed tracks according to the z -coordinate at the point of closest approach to the beam line. An adaptive vertex fit is used on these sets of tracks to find the position of the vertex [151]. The primary interaction vertex is defined as the primary vertex with the largest sum of p_T^2 from the objects associated to it. A dedicated algorithm is used to identify secondary displaced vertices within the tracker volume, using the complete set of reconstructed tracks [152, 153]. These displaced vertices mainly originate from nuclear interactions with the tracker material or from conversion photons.

Muon track elements are identified using two different methods. The first matches each track element present in the muon chambers to a track reconstructed in the inner tracking system and a track fit performed using a Kalman filter. These tracks are called global muon tracks. The second method takes all tracks in the tracker and looks for compatible signatures in the calorimeters and in the muon chambers. If a compatible signature is found then the inner track is labelled as a tracker muon track.

Calorimetric cluster candidates are constructed individually, but with a similar methodology, for the ECAL and HCAL. Initially, cluster seeds are created which are the local energy deposit in a cell which is maximal with respect to adjacent cells and above a seed energy threshold. Next, bordering cells are added to the seed if they contain a deposit above a given cell energy threshold. All thresholds can be found in [149]. Calorimeter clusters not associated to an extrapolation of a track candidate are clear indicators of a neutral particle. Care must be taken however, when a neutral-particle deposit overlaps a charged-particle deposit. These are detected as an excess in cluster energy with respect to the sum of the p_T of the associated charged particles. The method of obtaining an accurate calorimeter calibration in order to maximise the identification of the neutral clusters, while minimising the misreconstructed energy excesses and to get the correct energy scale for all neutral particles is also given in [149].

5.1.2 Linking algorithm

The next step in reconstructing a particle is to use a link algorithm to connect all the requisite PF elements together. The probability for the link algorithm to connect two elements together depends on the granularity of the CMS subdetectors and the number of particles to resolve. For a single particle, the probability to link all elements together depends predominantly on the volume of material traversed before the calorimeters due to nuclear interactions introducing trajectory kinks and secondary particles.

A link between the central track and calorimeter cluster is created if the extrapolation of the track candidate from the last hit position to a depth of one interaction length in the calorimeters is within a cluster candidate area. Neutral clusters in the ECAL are linked to electron bremsstrahlung radiation if a tangential extrapolation (starting from tracker hit position) lies within the boundaries of a cluster candidate and the η displacement between the cluster and original track candidate extrapolation is less than 0.05. The bremsstrahlung photons have a high chance to form a pair of conversion electrons (e^+e^-). A conversion finder is used to find links between compatible track candidates originating from a photon conversion and if found is subsequently linked to the originating track candidate. Links between ECAL, HCAL and preshower cluster candidates are formed if they are outside of tracker acceptance. The cluster position in the more granular subdetector must lie within the cluster envelope of the coarser subdetector. Charged particle tracks associated to a common secondary vertex are linked. Finally, links can be created between a central track candidate and information from the muon

detectors either as a global or a tracker muon track. All PF elements linked together form a PF block.

The PF algorithm uses the PF blocks as an input and performs reconstruction and identification in the following order. Muon candidates are identified and reconstructed before removing the corresponding PF elements from the PF block. Next, electrons are processed including bremsstrahlung radiation concurrently with isolated photons and again associated PF elements removed. Finally, the remaining PF elements are associated to charged hadrons, neutral hadrons and photons emanating from the fragmentation and hadronisation occurring within parton jets. The reconstruction and identification of each PF object is described in the following sections.

5.1.3 PF muons

PF muons are identified from the global and tracker muon track properties. First, isolation requirements are imposed on the global muon tracks. Isolation is a measure of the activity of the event around the object of interest. In this case, additional tracks and calorimeter deposits are taken within a cone of

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3, \quad (5.1)$$

and if

$$\sum p_T^{\text{track}} + \sum E_T^{\text{calo}} < 0.1 \times p_T^{\text{muon}}, \quad (5.2)$$

then no further selection is required as hadrons misidentified as muons are rejected adequately. Above this value, charged hadrons are susceptible to being misreconstructed as muons, due to punch-through from the back of the HCAL. In this regime, the global muons must pass a tight-muon identification criteria, described in Sec. 5.2.2, and either at least three matching hits found in the muon chambers or the calorimeter deposit is compatible with the muon hypothesis. If the muon does not pass the tight-muon identification (ID) due to a poor inner or global fit, then it is retained if the tracker muon fit is of high quality and the calorimeter deposit is compatible with the muon hypothesis.

For muons with $p_T < 200$ GeV, the muon track resolution is better for tracker muons than for global muons and so the tracker muon track is used to determine the muon properties. Otherwise they are taken from the global or tracker muon track fit with the minimum χ^2 .

5.1.4 PF electrons and photons

Due to the similarity in the basic properties of electrons and isolated photons (electrons emit bremsstrahlung photons and photons convert to e^+e^- pairs), the PF algorithm reconstructs isolated PF photons and PF electrons simultaneously. PF electrons are seeded from a cluster candidate linked to a Gaussian Sum Filter (GSF) track [154], given that it is not additionally linked to more than three other tracks. Isolated PF photons are seeded from clusters > 10 GeV with no link to a GSF track.

GSF tracks are based on a combination of two different methods of track reconstruction. The first is the ECAL-based method, which extrapolates to the inner track hits depending on the ECAL cluster energy and position. To seed the ECAL-based track all clusters of $E_T > 4$ GeV are used with the assumption that they are produced by either an electron or positron. A large proportion of the energy of the electron or positron is lost to bremsstrahlung and so the performance of the method depends strongly on the ability to collect all the associated energy depositions, while excluding depositions from other particles. To collect the energy, the cluster is extended in ϕ around the electron direction to account for the bending in the magnetic field. The group of clusters is known as a supercluster (SC). The ECAL-based track

seeds suffer at low- p_T due to the large bending in the magnetic field and in jets due to the non-isolated environment. The second is the tracker-based method and recovers electron tracks missed by the ECAL-based approach. It starts from the iterative-based track candidates with a $p_T > 2$ GeV, with more specific requirements of the number of missing hits and the χ^2 of the track fit. A reduced GSF fit is used on the track hits as it is more suitable to electron fitting than a Kalman filter. A final selection on a boosted-decision-tree [155] that combines the discriminating power of a number of distributions, e.g. the χ^2 of the GSF fit, the distance between the extrapolated track and the nearest cluster and the energy lost along the GSF track is used to select tracker-based track seeds. The track seeds obtained by the two methods are combined together and a final full GSF fit processed.

PF electrons are required to satisfy additional identification criteria on both the track seed and cluster properties. These criteria include the track fit χ^2 , the ratio of energy in the HCAL to the ECAL and the number of hits. A full set of parameters is given in [149]. PF photons are selected depending on compatibility with the photon hypothesis of the isolation with respect to GSF tracks and clusters, the shower shape and the ECAL to HCAL energy ratio.

5.1.5 PF jets

Once the electrons, isolated photons and muons have been reconstructed and removed from the PF blocks, the remaining particles must be from neutral hadrons, charged hadrons or non-isolated photons formed from the hadronisation of the final state. Hadronic jets are composed of these hadrons and photons, where the photon content deposits primarily in the ECAL, the neutral hadron content in the HCAL and charged hadron content in both calorimeters.

All calorimetric PF elements that are not linked to a tracking element are considered to be either neutral hadrons or photons. Within the tracking acceptance the charged and neutral hadrons can be identified separately and as the neutral hadrons only leave $\approx 3\%$ of the jet energy in the ECAL, the ECAL deposits are considered to originate solely from photons and the HCAL deposits from neutral hadrons. Outside the tracking acceptance however, there is no way to separate charged hadrons from neutral hadrons. As the charged hadrons also leave deposits in the ECAL, it is not possible to assign the deposits to photons alone. Now the links between ECAL deposits and HCAL deposits become important. ECAL deposits not linked to HCAL are considered to be photons, while linked deposits are considered to arise from a neutral or charged hadron.

Once the neutral hadrons and non-isolated photons have been identified and removed, there are only calorimeter elements linked to one or more track elements remaining. These make up the charged hadrons. PF jets are created by using the anti- k_T jet clustering algorithm [156] on the charged and neutral hadrons.

Anti- k_T jet clustering algorithm

The anti- k_T jet clustering algorithm is an extension to the Cambridge/Aachen (C/A) and k_T jet clustering algorithms. These algorithms work by calculating the distance, d_{ij} , between two entities, either particles or proto-jets, and the distance between an entity and the beam line, d_{iB} . If the smallest distance is between two entities then they are recombined into a new proto-jet, whereas if d_{iB} is smaller, then i is called a jet and removed from the list of entities. All distances are then recalculated and the clustering repeated until no entities remain. The maximum cone size is given by the radius parameter, R , which for jet reconstruction during 2016 data taking was set at $R = 0.4$. The differences between the anti- k_T and the (C/A) and k_T algorithms occurs in the definition of the distance parameters

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (5.3)$$

and

$$d_{iB} = k_{Ti}^{2p}. \quad (5.4)$$

The parameter p gives the strength of the relative contributions to d_{ij} and d_{iB} from the hardness of the entity and the angular separation. The angular separation is defined as

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}. \quad (5.5)$$

For $p = +1$ the k_T algorithm is recovered and the Cambridge/Aachen algorithm when $p = 0$. The anti- k_T algorithm uses $p = -1$ which means that precedence is given to the hardness of the entities, e.g. a soft particle is much more likely to cluster to a hard particle than another soft one. This leads to perfectly conical jets if no two initial hard particles are within $2R$ of each other. If there are two hard particles within $R < \Delta_{12} < 2R$ but $k_{T1}^2 \gg k_{T2}^2$ then the hardest jet will be conical and the softer a cone missing the overlap. If $k_{T1}^2 \sim k_{T2}^2$ then the jet boundary will lie between the two jet axes. For closely spaced hard particles within $\Delta_{12} < R$, clustering will occur, but for $k_{T1}^2 \sim k_{T2}^2$ the shape becomes a union of the cones of both constituent hard particles as well as the cone centred on the final jet. This means that the anti- k_T algorithm is resilient against soft radiation but adaptive to hard emission. An example anti- k_T jet clustering is shown in Fig. 32.

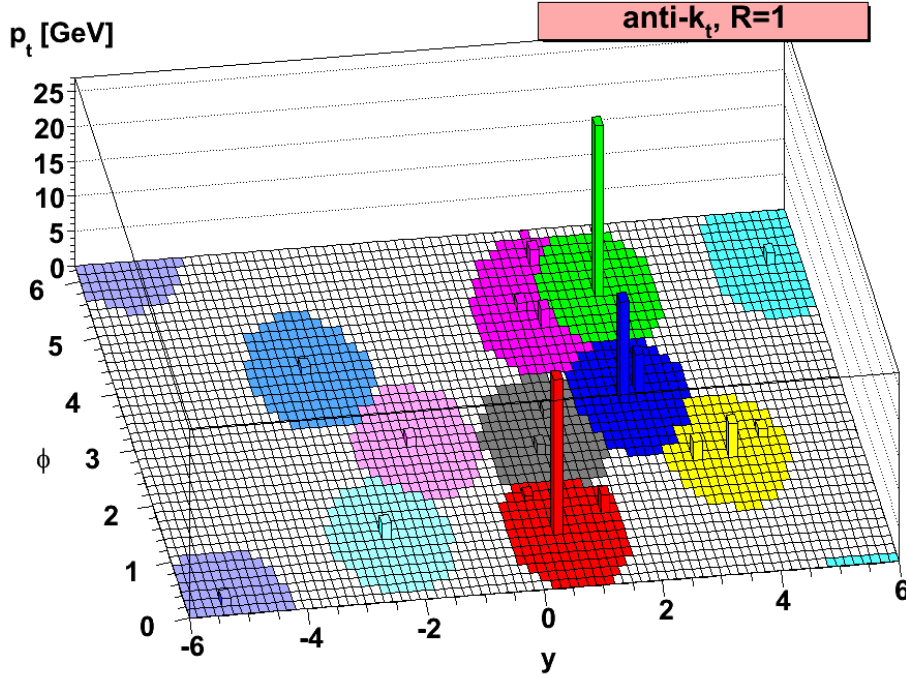


Figure 32: An example of clustering using the anti- k_T algorithm. Soft particles are preferentially clustered around high- p_T seeds forming conical jets. The jet boundary is defined by the respective hardness between neighbouring jets. Figure taken from [156].

As the particle flow algorithm attempts to reconstruct every individual particle in an event, before the jet clustering, it is possible to assign charged hadrons to either a high- p_T interaction vertex or an additional interaction vertex. The hadrons associated to additional interaction vertices are known as in-time pileup. If a charged jet constituent is matched to an additional interaction vertex, then it can be removed from the jet. This is known as charged hadron subtraction. If there is no clear association then the charged hadrons remain in the jet. Pileup subtraction is discussed further in Sec. 5.2.3.

5.1.6 PF missing transverse momentum

Conservation of momentum implies that the vector sum of all transverse momenta is zero in a head-on collision. The presence of a momentum imbalance, \vec{p}_T^{miss} , implies an undetectable particle in the event, e.g. a neutrino. In terms of particle flow, \vec{p}_T^{miss} is given by the negative vector sum of all the measured particle flow objects

$$\vec{p}_T^{\text{miss}} = - \sum_i^{\text{PF objects}} \vec{p}_T^i. \quad (5.6)$$

More generally, the magnitude of this missing momentum is used, p_T^{miss} . There is, however, a small probability that it is misreconstructed to be artificially large. The cause of this is most often the misreconstruction or misidentification of a high- p_T muon. If an event has a large p_T^{miss} then the PF algorithm applies a post processing step. The first possibility investigated are genuine muons from the cosmic ray muon background coincident with a bunch crossing. They are recognised as such if their trajectory is more than 1 cm away from the beam line and removed from the particle flow collection if the p_T^{miss} is reduced by at least one half. The second possibility relates to bad reconstruction of the momentum of the muon, indicated by significant differences in the global and tracker muon p_T estimates. The bad reconstruction can be caused by interactions in the steel return yoke, in-flight decays and incorrectly assigned hits in the inner tracker. If the p_T^{miss} is reduced by at least one half when recalculating with respect to the different muon p_T estimates then the muon estimate with the lowest p_T^{miss} is taken. The third possibility comes from the misidentification of a charged hadron punching through into the muon chambers. A muon is reconstructed together with an energetic neutral hadron from the calorimeter deposits. Finally, another, similar possibility is that of an overlapping muon and neutral hadron deposit being misreconstructed as a charged hadron deposit. The particles are substituted with each other respectively if the p_T^{miss} is reduced by at least one half.

The p_T^{miss} is corrected accordingly when jet energy corrections (JEC), explained in Sec. 5.2.3, are applied to the PF jets.

$$\vec{p}_T^{\text{miss}} = - \sum_i^{\text{PF objects}} \vec{p}_T^i - \sum_j^{\text{PF jets}} (\vec{p}_T^{j, \text{corr}} - \vec{p}_T^j). \quad (5.7)$$

5.1.7 PF isolation

The requirement of leptons to be isolated has already been used in the reconstruction of the PF objects. Often, more stringent isolation criteria are required and so the PF relative isolation, I_{rel} , is employed. It is defined as the ratio of the additional p_T within a cone of set distance around the lepton to the p_T carried by the lepton. The additional p_T includes all charged and neutral hadrons as well as photons present within the cone.

$$I_{\text{rel}}^\ell = \frac{\sum p_T^{\text{charged}} + \sum p_T^{\text{neutral}} + \sum p_T^\gamma}{p_T^\ell}. \quad (5.8)$$

In addition, contributions from additional interactions can be taken into account. Charged particles from these extra interactions are not associated to the primary vertex and are removed. Direct removal is not possible for neutral particles produced by the pileup. One possible correction for this neutral pileup is the $\Delta\beta$ correction. This uses the ratio of charged hadrons to neutral hadrons and photons in inelastic collisions to estimate the contribution of non-charged pileup from charged pileup. The correction $\Delta\beta$ is set to 0.5 [157]. The relative isolation then

becomes

$$I_{\text{rel}}^{\ell} = \frac{\sum p_{\text{T}}^{\text{charged}} + \max(0, \sum p_{\text{T}}^{\text{neutral}} + \sum p_{\text{T}}^{\gamma} - \Delta\beta \sum_{\text{pileup}} p_{\text{T}}^{\text{charged}})}{p_{\text{T}}^{\ell}}. \quad (5.9)$$

5.2 Analysis object selection

While the PF algorithm is good at resolving particles, it is sometimes necessary to impose stricter identification criteria. By increasing the probability that an analysis object is correctly reconstructed, the selection efficiency for that object decreases correspondingly. When the statistical uncertainty is not an issue or where an analysis requires a high object purity a tighter selection is usually used. This is the case for $t\bar{t}$ events produced at the LHC, having both a high cross section and large data set sizes. In the following the identification criteria for all the objects used in this thesis will be detailed.

5.2.1 Electron objects

As stated in Sec. 3.3.8, events that pass the HLT_Ele32_eta2p1_WPTight_Gsf trigger are selected as the basis of the e+jets channel in this thesis. A harder p_{T}^{ℓ} requirement is chosen at $p_{\text{T}}^{\ell} > 34$ GeV to operate in a region optimal with respect to the trigger efficiency. The $|\eta^{\ell}| < 2.1$ requirement is maintained, however electrons falling into the gap between the barrel and endcap ($1.444 < |\eta^{\ell}| < 1.566$) are discarded. The signal electron object must pass a set of strict identification requirements (tight ID). An additional collection of veto electrons are selected using a lower $p_{\text{T}}^{\ell} > 15$ GeV and a looser set of identification criteria (veto ID).

Electron identification

Further identification requirements are placed on the PF electrons in order to be used as analysis objects. The additional requirements are combined into an electron identification which provides more discrimination for isolated signal electrons against hadronic jets and nonprompt sources of electrons. The electron identification criteria are defined in [158] and the tight and veto IDs applied are shown in Tab. 6 for both the barrel and endcap regions of the detector. The average efficiency of the tight ID is designed to be 70% and the veto ID to be 95%.

The shower shape variable $\sigma_{i\eta i\eta}$ is the standard deviation of the position of the energy deposited in a 5×5 lattice centred in the seed crystal. An electromagnetic shower will be much more confined than a hadronic shower and therefore have a lower value of $\sigma_{i\eta i\eta}$. The absolute difference in η between the SC position and the electron track seed extrapolated to the ECAL assuming no bremsstrahlung radiation is given by $|\Delta\eta_{\text{seed}}|$. Similarly, the absolute difference in ϕ between the SC position and the track direction at the primary interaction vertex extrapolated to the ECAL assuming no bremsstrahlung radiation is given by $|\Delta\phi_{\text{in}}|$. Additional particles present in hadronisation processes bias the position of the SC due to the increased bending of low- p_{T} radiations. The ratio of the hadronic energy to the electromagnetic energy H/E is smaller for deposits from electrons than from hadrons. A tighter requirement on the PF isolation, with effective area pileup correction, using a cone size of $R = 0.3$, is included. The effective area method removes pileup contributions by calculating the product of the effective area that the electron is taking up in the detector and the median energy density of the event. The median energy density is also known as the offset energy density and is discussed in more detail in Sec. 5.2.3. It is more suitable than the $\Delta\beta$ correction, introduced in Sec. 5.1.7, due to being insensitive to the electron radiations. The variable $|1/E - 1/p|$ is proportional to the radius of curvature of the electron track and gives additional discrimination against electrons from conversion photons. The missing inner hits requirement provides a stricter track quality

acceptance than present in the PF algorithm. A selected electron is not allowed to originate from an identified conversion photon. In addition to the electron identification, the electron must originate close to the primary vertex d_0 and close to the beam line d_z .

Table 6: The identification criteria for signal electrons (tight) and veto electrons (veto). The criteria are split into the barrel and endcap regions at $|\eta_{SC}| = 1.479$

ID variable		Tight		Veto	
		Barrel	Endcap	Barrel	Endcap
full $5 \times 5 \sigma_{in\eta}$	$<$	0.00998	0.0292	0.0115	0.037
$ \Delta\eta_{seed} $	$<$	0.00308	0.00605	0.00749	0.00895
$ \Delta\phi_{in} $	$<$	0.0816	0.0394	0.228	0.213
$\frac{H}{E}$	$<$	0.0414	0.0641	0.356	0.211
PF I_{rel} with EA correction	$<$	0.0588	0.0571	0.175	0.159
$ \frac{1}{E} - \frac{1}{p} $	$<$	0.0129	0.0129	0.299	0.15
Missing Inner Hits	\leq	1	1	2	3
Conversion Veto		Y	Y	Y	Y
Additional to ID					
d_0 (cm)	$<$	0.05	0.10	0.05	0.10
d_z (cm)	$<$	0.10	0.20	0.10	0.20

5.2.2 Muon objects

Signal analysis muons must have a $p_T^\ell > 26$ GeV, similarly imposed from the efficient operation region of the HLT_IsoMu24 \cup HLT_IsoTkMu24 trigger selection and must pass tight ID requirements. Veto analysis muons have a $p_T^\ell > 15$ GeV, aligned with that in the e+jets channel and must pass a loose muon ID.

Muon identification

Analysis muons are selected from the PF muons by applying a muon identification in an identical manner to the analysis electrons. The criteria of the identity are listed in Tab. 7 [159]. Similarly to the electrons, the tight ID is designed to operate at an average efficiency of 70%, and the loose ID at 90%.

For signal muons, hits are required along the entire path of the muon from the silicon pixel detector to the muon chambers. A good χ^2/ndf of the track fit is also required. Constraints are placed with respect to the primary vertex and beam line as with the electron identification. Finally, a more stringent selection on the PF I_{rel} using the $\Delta\beta$ pileup correction using a cone size of $\Delta R < 0.4$ is required.

5.2.3 Jet Objects

On top of the charged hadron subtraction in the PF algorithm, analysis jets are required to have $p_T > 30$ GeV, $|\eta| < 2.4$ and pass a loose ID shown in Tab. 8. The jet ID for PF jets was introduced in [160]. To ensure a jet is not additionally classified as a lepton it must have a minimum distance to the nearest lepton $\Delta R > 0.4$. If the jet is close to a lepton then it is assumed to have originated from that lepton and is consequently removed from the jet collection.

Table 7: The identification criteria for signal muons (tight) and veto muons (loose).

ID variable		Tight	Loose
is PF muon		Y	Y
is global or tracker muon		Y	Y
Pixel hits	>	0	—
Tracker layer hits	>	5	—
Muon stations hit	>	2	—
Muon chambers hit	>	1	—
χ^2/ndf	<	10	—
d_0 (cm)	<	0.2	—
d_z (cm)	<	0.5	—
Additional to ID			
PF I_{rel} with $\Delta\beta$ correction	<	0.15	0.25

Table 8: The identification criteria for signal jets.

ID variable		Loose
Neutral hadron fraction	<	0.99
Charged hadron fraction	>	0
Neutral electromagnetic fraction	<	0.99
Charged electromagnetic fraction	<	0.99
Number of constituents	>	1
Number of charged constituents	>	0
Muon fraction		—

The response of the reconstructed jets is not equal over the entire phase space of the CMS experiment and needs to be corrected. These are called jet energy corrections (JECs) and consist of correction to the jet energy scale (JES) and jet energy resolution (JER). Once the reconstructed jets have been fully calibrated, it is possible to probe them to try and identify whether it has been initiated from a b quark. These calibrations and indentifications are discussed in the following sections.

Jet energy corrections

The response from the detector calorimeters to particle energy is not linear meaning that a simple translation between the reconstructed jet energy and the calibrated jet energy is not feasible. JECs are used to provide the mapping between the measured and calibrated jet energies. The JECs are applied sequentially with each step correcting the jet energy scale (JES) according to the associated jet parameters (p_T , η , flavour). These steps are shown in Fig. 33.

The first correction is applied to counteract additional energy in the jet, known as offset energy, coming from pileup, the underlying event and electronic noise. In-time pileup is already partially removed by the charged hadron subtraction algorithm, which accounts for $\sim 50\%$ of the offset energy. The rest of the diffuse neutral energy and out-of-time pileup is removed on an event-by-event basis using the hybrid jet area method dependent on the offset energy density, ρ , and the effective jet area, A^{jet} . A^{jet} is estimated by artificially adding a very large number of infinitely soft four-momenta to the event, so as to have no effect on the properties of the true jet, and clustering them into the true jets. The extent in $\eta - \phi$ is the active jet area. The quantity ρ is taken as the median value of $p_T^{\text{jet}}/A^{\text{jet}}$ for all jets in the event. The use of the median

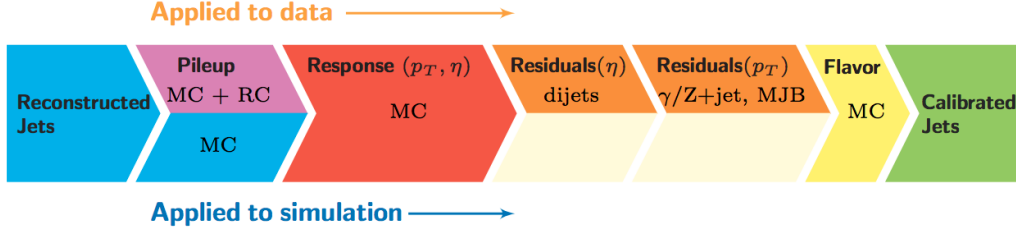


Figure 33: A schematic showing the factorised steps to correct the energy scale of a jet. These consist of a pileup correction, detector response correction with additional residuals applied in data and optional flavour corrections. All corrections marked with MC are derived from simulation studies, RC stands for random cone, and MJB refers to the analysis of multijet events. Schematic taken from [161].

creates an insensitivity to hard jets in the event. The correction is applied as

$$p_T^{\text{corr}} = p_T^{\text{uncorr}} - A^{\text{jet}} \left(\rho_0(\eta) + \rho\beta(\eta)[1 + \gamma(\eta) \log(p_T^{\text{uncorr}})] \right), \quad (5.10)$$

where ρ_0, β and γ are η -dependent variables which are estimated from simulated QCD dijets processed with and without pileup overlay and parametrise the shape of the offset energy distribution. Residual corrections are applied to data to correct for any differences seen in the offset energy distribution seen in the data, using the random cone method on zero-bias events, and simulation. Zero-bias events are events in which there is no hard-scattering process and so any jet energies should come from pileup provided the electronic noise is small. The random cone method estimates ρ by clustering jets around random cones, which map the full $\eta - \phi$ space, and taking the average jet p_T . This correction removes any dependence of the jet energy on the luminosity for the subsequent corrections.

Next, a second set of corrections are applied to both data and simulation to create a uniform jet energy response with respect to p_T and η . This is done using a simulated QCD dijet sample to compare the p_T of the reconstructed jets to the particle-level jets. The response is defined as the ratio of the average reconstructed jet p_T to the average particle-level jet p_T . The left panel of Fig. 34 shows the simulated jet response at the end of Run 1 [161], which is used in the JEC applied in this thesis. The preliminary measurement of the simulated jet energy response using detector conditions during 2016 [157] is shown in the right panel of Fig. 34. The comparison shows that the response of the jets has not changed significantly between different Run periods, especially when considering jets with respect to the object selection used, $p_T > 30$ GeV and $|\eta| < 2.4$.

Residual corrections are calculated from data and applied to correct small differences seen in the jet energy response between simulation and data. These consist of a η -dependent correction determined from QCD dijet events relative to a barrel reference jet of similar p_T and a correction on jet energy scale with respect to p_T . The p_T correction is based on Z/γ +jets and QCD multijet events. Finally, optional corrections based on the response of different particle flavours can be applied. These flavour corrections are derived from simulation and are not applied in this thesis.

The JECs correct the JER as well as the JES. The JER scale factors to do this, SF_{JER} , are determined from the ratio of data to simulated QCD dijet events with respect to p_T, η and the number of additional interaction vertices. On top of aligning the average values of the JER, the spread of the resolution, which in simulation is smaller than that seen in data, needs to be smeared in simulation to match data. A hybrid of two approaches is used to correct and smear the resolution in simulation. If a calibrated jet is well matched to a particle-level jet such that the distance between the centres of the cones of each jet is less than half the cone size

$$\Delta R < \frac{R}{2}, \quad (5.11)$$

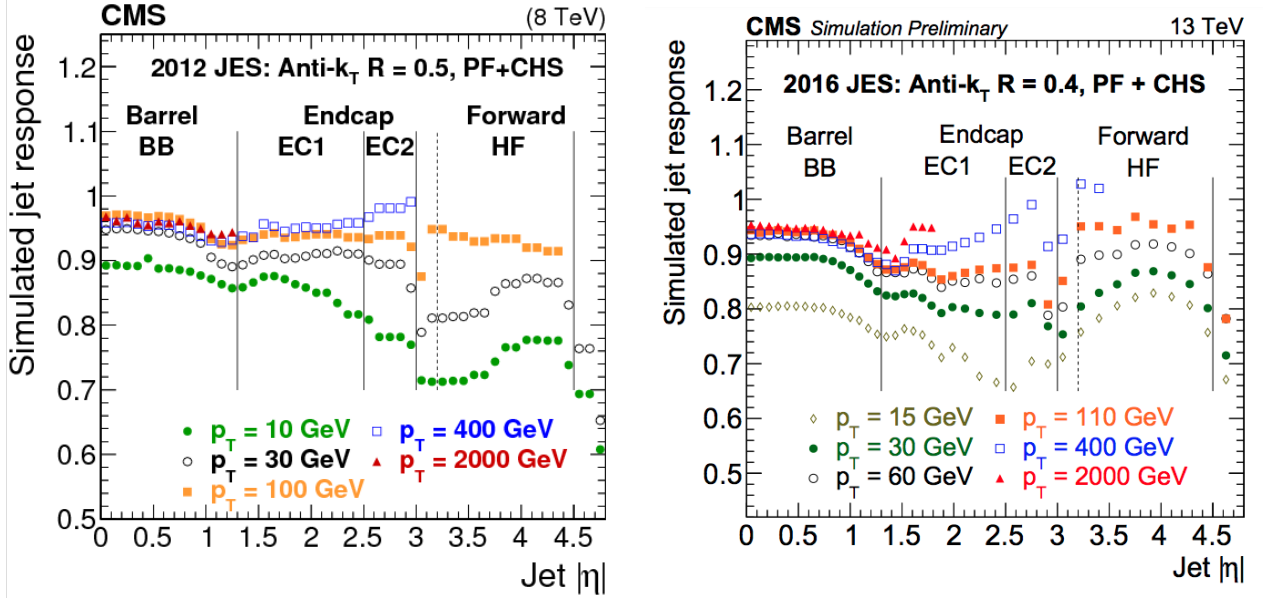


Figure 34: The simulated jet energy responses at 8 TeV, in the left panel, and 13 TeV, in the right panel, as a function of p_T and η . Figures taken from [161] and [157] respectively.

and the difference in p_T satisfies

$$|p_T - p_T^{\text{ptcl}}| < 3 \sigma_{\text{JER}}^{\text{sim}} p_T, \quad (5.12)$$

where $\sigma_{\text{JER}}^{\text{sim}}$ is the relative p_T resolution as measured in simulation (σ_{p_T}/p_T), then the jet p_T is rescaled by

$$1 + (\text{SF}_{\text{JER}} - 1) \frac{p_T - p_T^{\text{ptcl}}}{p_T}. \quad (5.13)$$

However, if the calibrated jet is not able to be well-matched then a stochastic smearing can be applied to degrade the resolution of the jet according to

$$1 + \mathcal{N}(0, \sigma_{\text{JER}}^{\text{sim}}) \sqrt{\max(\text{SF}_{\text{JER}}^2 - 1, 0)}. \quad (5.14)$$

The essentially smears the SF_{JER} to be applied to each jet by multiplying it with a random number sampled from a Gaussian distribution, $\mathcal{N}(0, \sigma_{\text{JER}}^{\text{sim}})$, centred on 0 with a standard deviation given by the relative p_T resolution in simulation. Both methods, ensure that the JER correction is never negative, instead being truncated at 0.

Identification of b quark jets

The properties of a jet arising from a b quark are different enough to those of a jet initiated from a light u, d or s quark or a gluon, that it is possible to discriminate between them. The ability to identify jets as originating from a b quark is very important for many areas of particle physics, especially those involving the top quark. This is because the top quark decays almost exclusively to a b quark and therefore a significant discrimination between signal top quark events and background processes can be achieved by adding a requirement on the number of b quark jets present in an event. The crucial aspect of the discrimination arises from the formation of B hadrons from the b parton, which have a longer lifetime and typically travel a few millimetres in the detector before decaying. The discriminating variables include the displaced tracks from the secondary vertex, the impact parameter, the increased fragmentation due to the higher b quark mass which can lead to both increased p_T measurements with respect

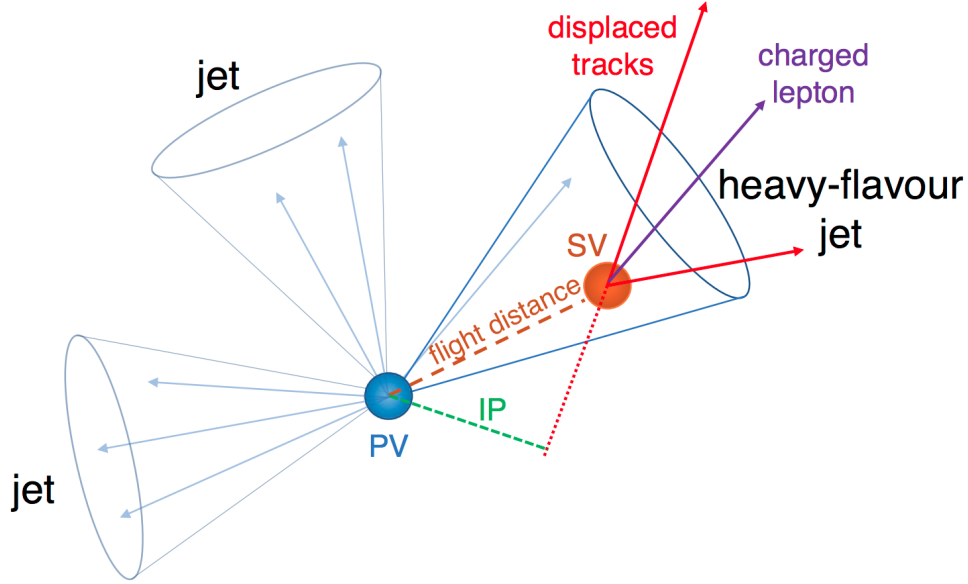


Figure 35: An illustration of a jet from a heavy flavour quark. A hadron is formed with a lifetime of ≈ 1 ps which decays at a displaced secondary vertex (SV) with respect to the primary interaction vertex (PV) and hence has a large impact parameter (IP). Figure taken from [162].

to the jet axis and a greater proportion of charged leptons in the jet constituents. Figure 35 shows an illustration of a heavy-flavour jet associated to a secondary vertex.

Several b quark tagging algorithms are documented in [162], which use multivariate techniques or deep learning to perform b quark jet discrimination. The combined secondary vertex algorithm (CSVv2), is chosen for this thesis because it is a well established, efficient algorithm. It assigns a value to each jet between 0 and 1 and where the closer the discriminating value is to 1, the more likely the jet is to originate from a b quark. The flavour composition of jets with respect to CSVv2 discriminant is shown in Fig. 36, after a basic event selection has been applied to a $t\bar{t}$ simulation. This event selection includes at least two jets with a CSVv2 > 0.8484 , defined as the medium working point of the algorithm, which is designed to have a misidentification rate of $\approx 1\%$. The b quark jets in this thesis also use the medium working point and a breakdown of the tagging efficiencies is shown in Sec. 6.3.

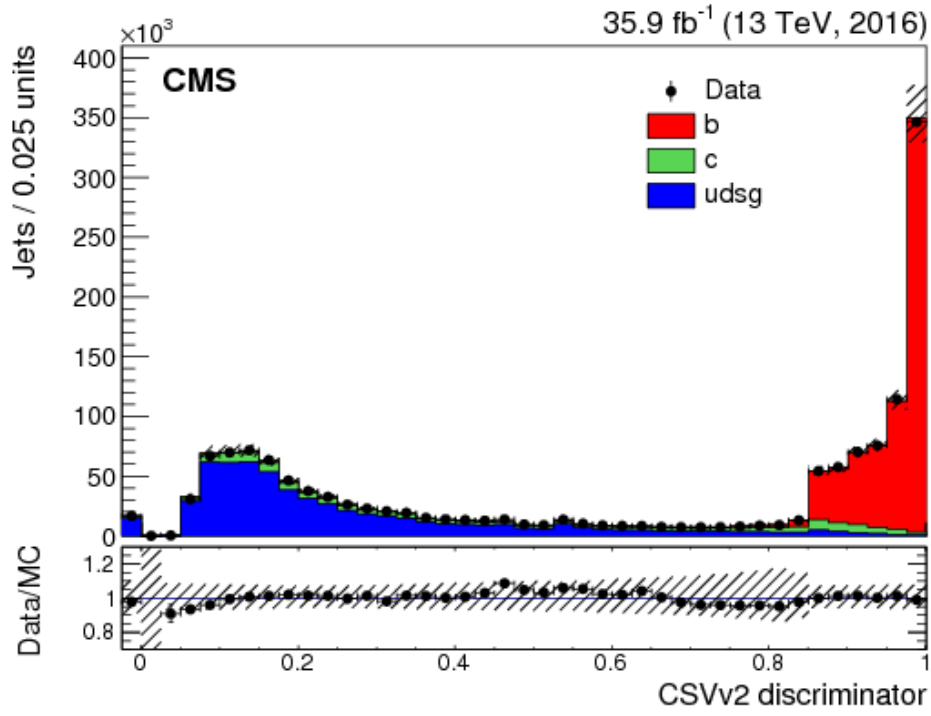


Figure 36: The flavour composition of jets from a $t\bar{t}$ model simulation. A simple event selection is performed requiring an isolated lepton, four jets with $p_T > 30$ GeV of which two are required to have $\text{CSVv2} > 0.8484$. Figure taken from [162].

CHAPTER 6

Event selection and corrections

With the complete reconstruction of events in both data and simulation it is possible to begin forming an analysis. An event selection, based on the single lepton final state of $t\bar{t}$ decay, can be created and applied using the analysis objects. The event yields are calculated as a function of the seven kinematic event variables to be used in the differential cross section measurements. Differences between data and simulation are observed in the efficiencies for the various algorithms applied, such as in the reconstruction or identification of leptons and so the yields in simulation need to be corrected to match data. Once the yields have been calculated after all corrections are applied, the reconstructed $t\bar{t}$ yield can be measured by subtracting the yields for the single top quark and vector boson backgrounds, as well as the data-driven estimate of the multijet-QCD background.

6.1 The event selection and kinematic event variables

The differential $t\bar{t}$ cross section measurements with respect to kinematic event variables is performed in the single lepton decay channel. The event selection is a set of criteria designed to reflect this final state and is applied on the collections of particles created following the methods described in Sec. 5.2. The event selection requires exactly one, isolated lepton, either electron or muon, with zero additional leptons. A minimum of four jets is needed, one from each of the $t \rightarrow bW$ decays, which is tagged as originating from a b quark and two from the hadronic decay of one of the W bosons. No upper limit is set on the jet multiplicity to allow for additional jets from initial and final state radiation to be present. An amount of missing p_T is expected in the event, stemming from the neutrino produced, however this is not included as part of the event selection. This is because the purity of the $t\bar{t}$ events is not improved significantly and it allows for the full p_T^{miss} spectrum to be used as one of the kinematic event variables of the measurement.

As stated in Sec. 2.7.3, the measurement described in this thesis relies on seven event variables which are N_{jets} , H_T , S_T , p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$. They are defined in the following way:

- N_{jets} is the multiplicity of jets in the event with $p_T > 30$ GeV and $|\eta| < 2.4$.
- H_T is the hadronic activity of the event and defined as scalar sum of the p_T of these jets.

- p_T^{miss} is defined as the magnitude of the missing \vec{p}_T , as defined in Sec. 5.1.6.
- p_T^ℓ is the magnitude of the \vec{p}_T of the signal lepton.
- $|\eta^\ell|$ is the absolute value of the pseudorapidity of the signal lepton.
- S_T is the total activity of the event, including contributions from the jets, signal lepton and p_T^{miss} . It is defined as $H_T + p_T^\ell + p_T^{\text{miss}}$.
- p_T^W is the magnitude of the \vec{p}_T of the leptonically decaying boson defined as $\sqrt{(p_x^\ell + p_x^{\text{miss}})^2 + (p_y^\ell + p_y^{\text{miss}})^2}$, where x and y are the Cartesian components of the transverse plane.

The event selection provides the reconstructed yield of events in data and simulation. There are differences present in the efficiencies of several algorithms that have been applied to data and simulation. The yields in simulation are subsequently reweighted to remove these differences in efficiency. The following sections describe the weights to be applied to the simulation.

6.2 Reweighting from the number of additional interactions

The number of additional proton-proton interactions in simulation differs from that seen in data. The simulation is corrected, per event, to match data by applying a scale factor. The scale factor is derived from the ratio of the expected number of additional interactions seen in data and simulation. The expected number of additional interactions seen in data, $N_{\text{PU}}^{\text{Data}}$, is estimated every bunch crossing using

$$N_{\text{PU}}^{\text{Data}} = \sigma_{\text{minbias}} \times L_i. \quad (6.1)$$

The parameter σ_{minbias} is the total inelastic cross section, taken as 69.2 mb, and L_i the instantaneous luminosity in the luminosity section i . The distribution of the number of additional interactions in simulation, $N_{\text{PU}}^{\text{Sim}}$, is taken from the expected probability distribution of the number of additional interactions, including interactions from out-of-time pileup. This is because the simulation is often generated before the running conditions in the detector are known. The scale factor applied to simulation is then calculated by

$$w_{\text{PU}} = \frac{N_{\text{PU}}^{\text{Data}}}{N_{\text{PU}}^{\text{Sim}}}. \quad (6.2)$$

As the number of additional interactions in an event can only be estimated in data, the reweighting can most easily be shown in an observable variable highly-correlated to it, that of the number of reconstructed primary vertices. The left panel of Fig. 37 shows the difference seen in the number of reconstructed primary vertices after full event selection, between simulation and data. The right panel of Fig. 37 shows the effect of applying the reweighting. It manages to correct some of the difference seen in the vertex multiplicity distribution, but not all. When varying σ_{minbias} up and down by its uncertainty of $\pm 4.6\%$, as seen in Fig. 38, the difference is covered. Therefore, it is reasonable to use the current additional interaction reweighting, when propagating the uncertainty through to the final measurements.

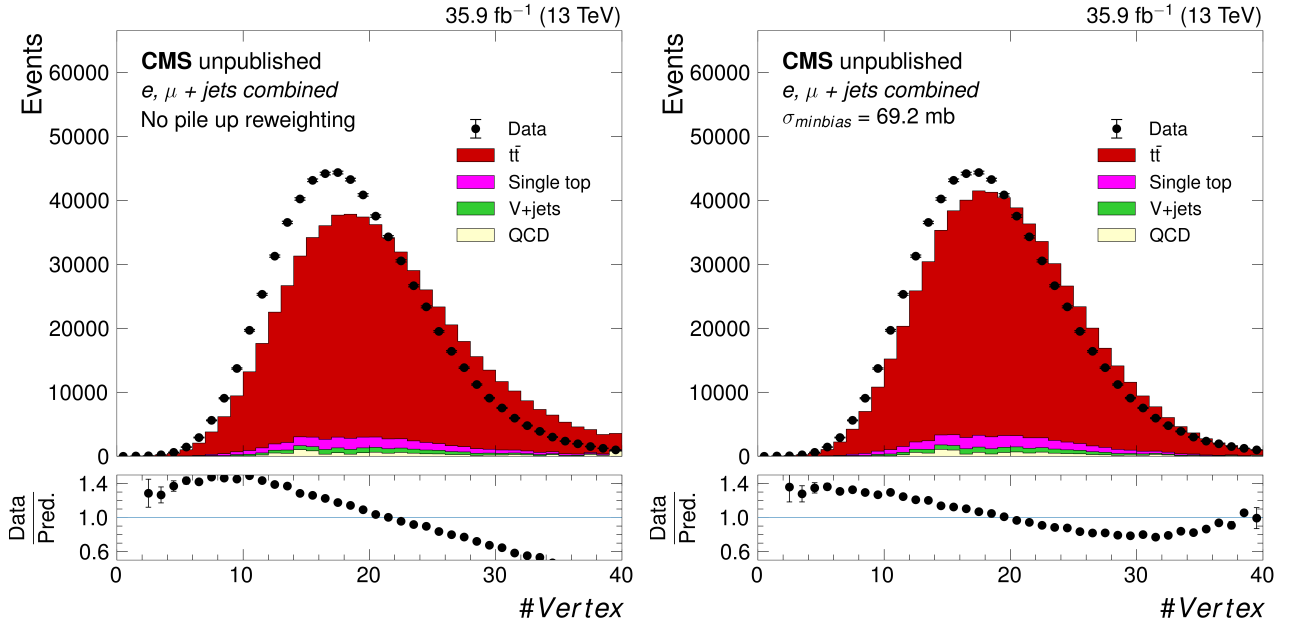


Figure 37: A plot showing the distribution of the number of reconstructed primary vertices in simulation and data. The left panel shows the distribution before pileup reweighting and the right panel after.

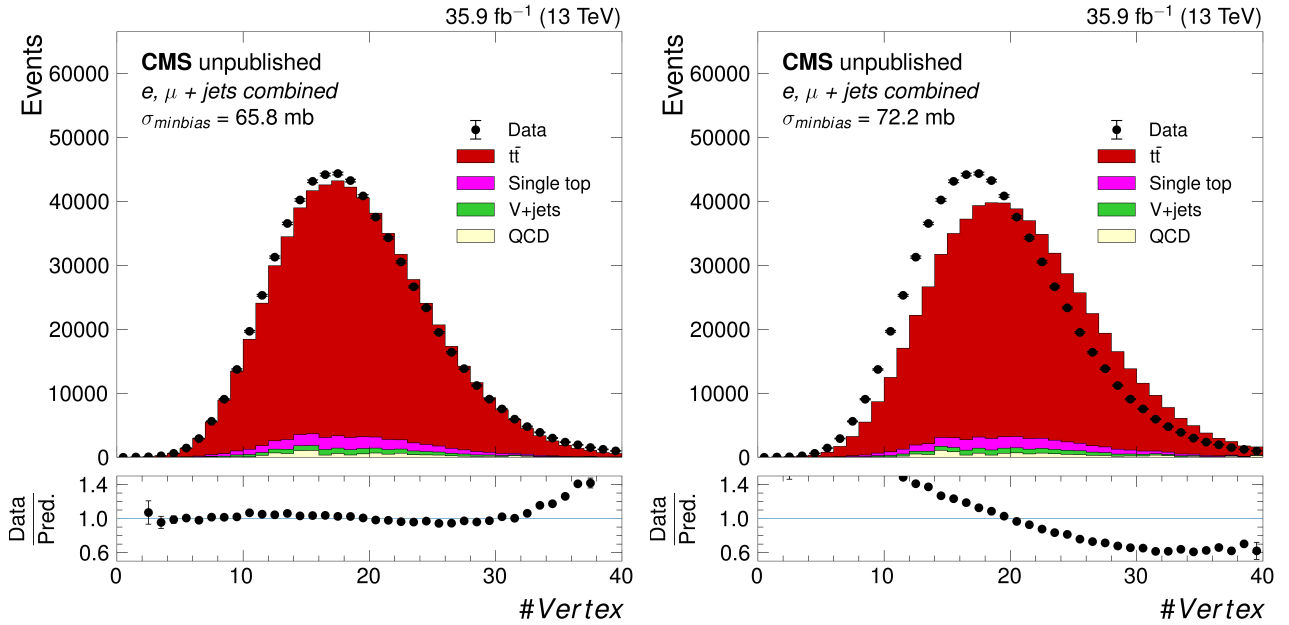


Figure 38: A plot showing the distribution of the number of reconstructed primary vertices in simulation and data. The left panel shows the reweighting after varying σ_{minbias} down by its uncertainty and the right panel after varying σ_{minbias} up by its uncertainty.

6.3 Reweighting from b jet identification

The probability for an event to contain a given configuration of i b tagged jets and j non-b tagged jets is not the same in data and simulation. The simulation is corrected, per event, to match data by applying a weight

$$w_b = \frac{P(\text{Data})}{P(\text{Sim})}, \quad (6.3)$$

where

$$P(\text{Sim}) = \prod_{i = \text{tagged}} \epsilon_i \prod_{j = \text{not tagged}} (1 - \epsilon_j) \quad (6.4)$$

and

$$P(\text{Data}) = \prod_{i = \text{tagged}} \text{SF}_i \epsilon_i \prod_{j = \text{not tagged}} (1 - \text{SF}_j \epsilon_j). \quad (6.5)$$

The scale factors $\text{SF}_i(p_T, \eta, \text{flavour})$ take into account the efficiency differences in b tagging between data and simulation ($\epsilon_{\text{DATA}}/\epsilon_{\text{MC}}$) and are derived as a function of p_T , η and flavour (b quarks, c quarks, or light quarks and gluons [udsg]). The b tagging scale factors are computed in [162], with samples of jets enriched in b quark content, such as dilepton $t\bar{t}$ or muon-enriched multijet events, and using a combination of methods such as likelihood fitting and the tag-and-probe method. The b jet tagging efficiency, $\epsilon_i(p_T, \eta, \text{flavour})$, is the efficiency for a jet to be b tagged in simulation and are also calculated as a function of p_T , η and flavour. They are derived for the CSVv2 algorithm, using the POWHEG+PYTHIA $t\bar{t}$ simulation and are taken as the ratio of the number of jets tagged as originating from a b quark to the total number of actual b jets. Figure 39 shows the b jet tagging efficiencies, which for genuine b jets is $\approx 60 - 70\%$ and for misidentified b jets stemming from c quarks is $\approx 15\%$, or from light quarks and gluons is $\approx 1\%$.

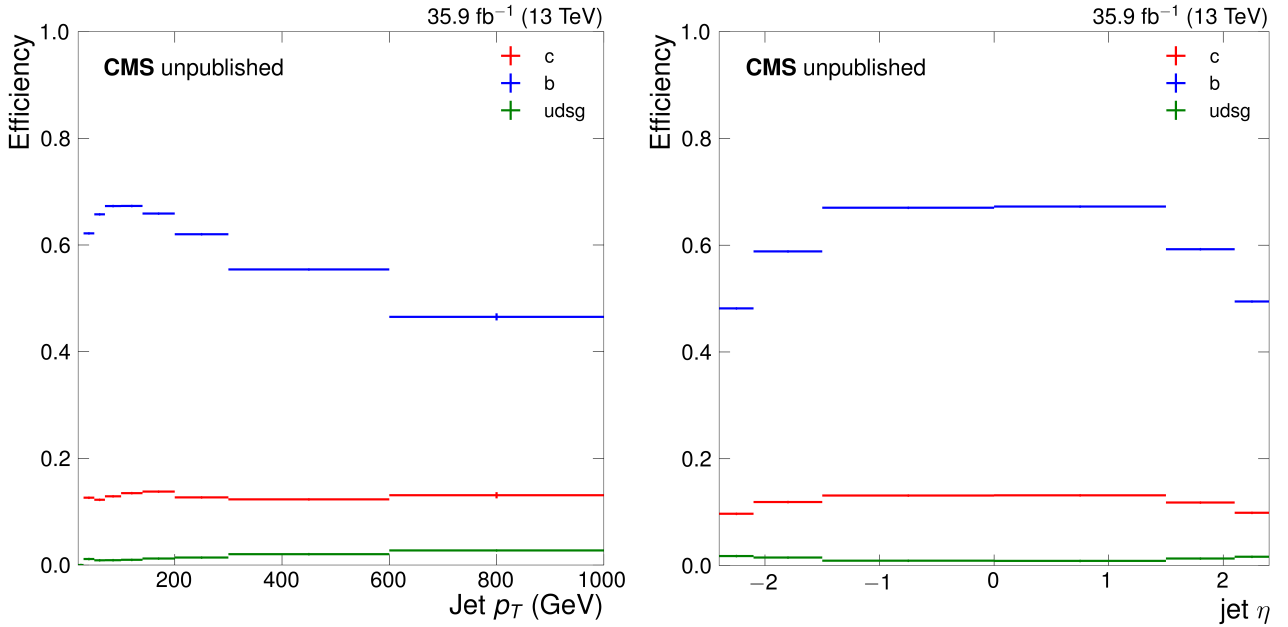


Figure 39: The left panel shows the b tagging efficiency with respect to jet p_T and the right panel with respect to jet η . Both panels show the efficiencies separated into b, c and light flavours.

The left panel of Fig. 40 shows the b jet multiplicity after the event selection, excluding the requirement on the multiplicity of b jets, before applying the b jet reweighting and the right panel after reweighting. The discrepancy in the first bin is covered by the uncertainty in

multijet QCD, taken from simulation. The size of the multijet QCD normalisation uncertainty is $\approx 60\%$, as motivated in Sec. 6.5.

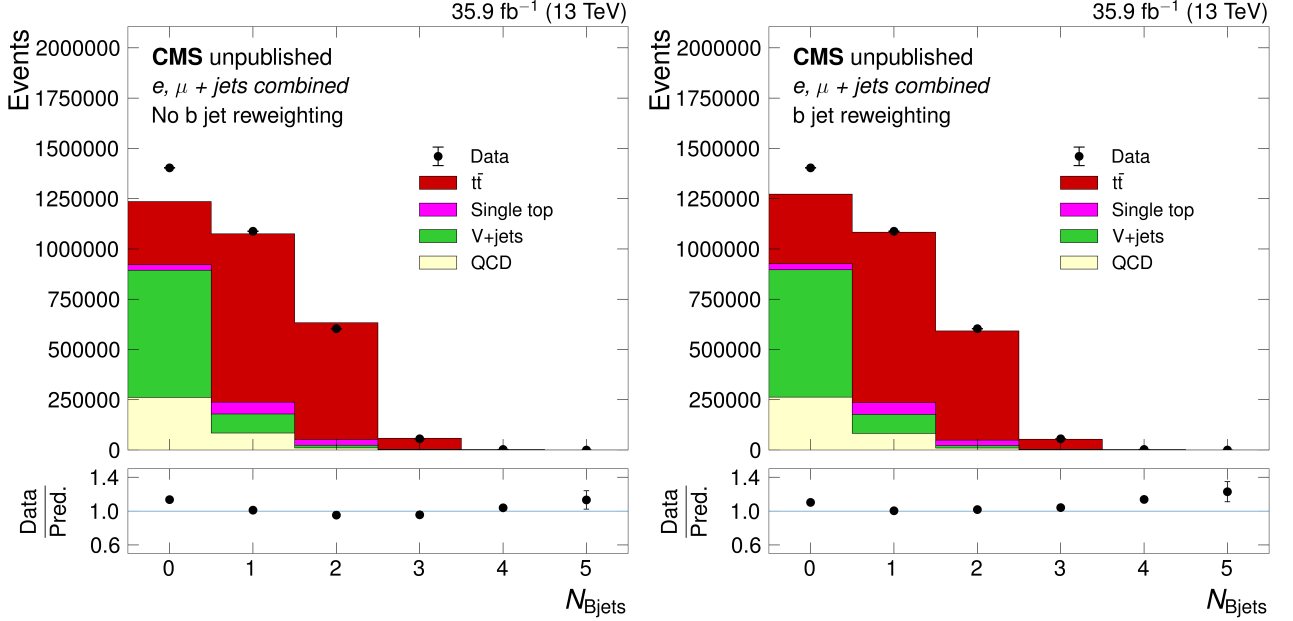


Figure 40: The left panel shows the b jet multiplicity distribution before reweighting and the right panel after reweighting.

6.4 Reweighting from leptons

The reconstruction, identification and triggering efficiencies of leptons are measured in data and simulation, and corrected in simulation to match data. The electron efficiencies and scale factors are derived using the tag-and-probe method following prescriptions defined in [158] using the clustering algorithm described in [149] and are presented in [163]. Muon scale factors are measured in a similar manner and are presented in [159, 164]. In both channels the efficiencies are measured from events containing a Z boson and the corrective scale factors are applied as a function of lepton p_T and supercluster pseudorapidity, η_{SC} . The total lepton scale factors vary between 0.95 and 1.

The electron ID scale factors are produced in bins of η_{SC} coarse enough to have a significant negative impact on measurements of $|\eta^\ell|$. This is because the high resolution of $|\eta^\ell|$ allows for fine-bin measurements and the application of the scale factors would cover several bins. It is therefore not appropriate to use these electron ID scale factors for the $|\eta^\ell|$ event variable. The scale factors are therefore rederived with respect to a finer η_{SC} binning scheme, which was used in the production of the electron reconstruction scale factors, over an identical p_T range. The rederivation used an identical method to that used to calculate the original scale factors. Figure 41 shows the efficiency of the electron ID in data (points) and simulation (dashes) and the scale factor to be applied is the ratio. This new scale factor is only applied for the $|\eta^\ell|$ differential measurements.

6.5 Multijet QCD

The multijet QCD background is notoriously hard to model and as such a data driven approach is used to estimate its contribution in the signal region from an orthogonal control region (often

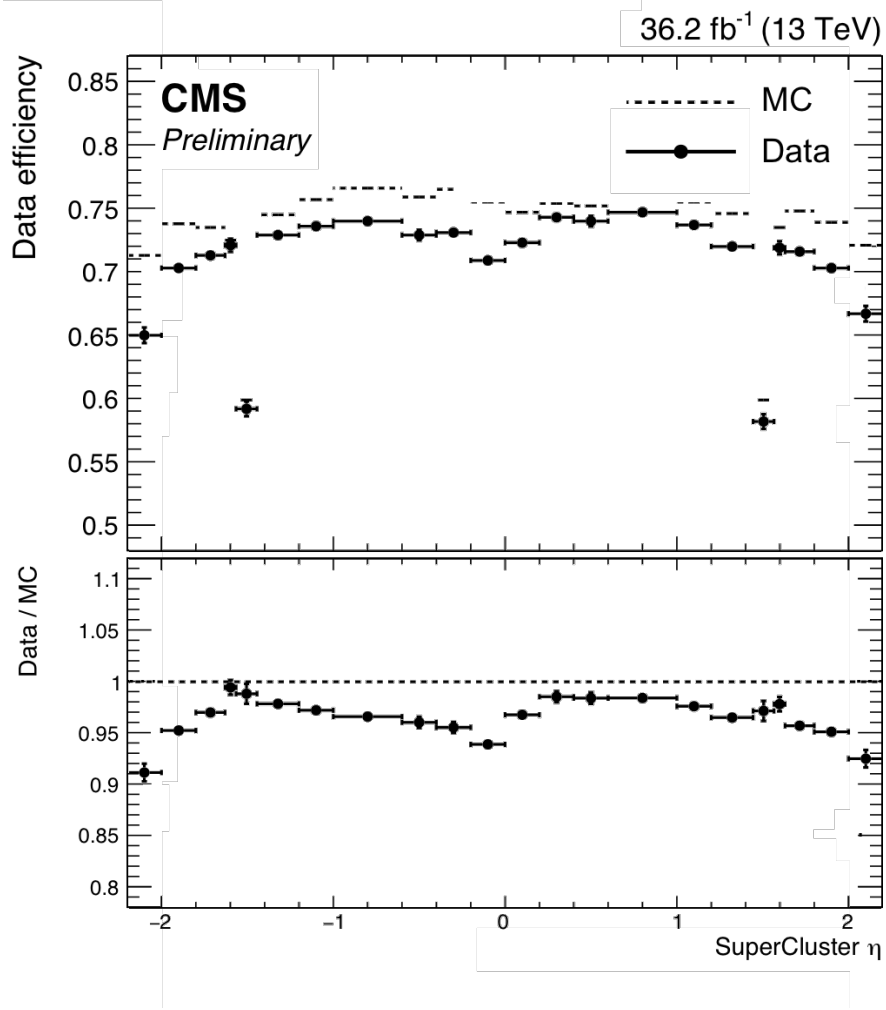


Figure 41: The electron identification efficiency for data and simulation calculated with events with a Z boson present using the tag-and-probe method. Electrons are selected within a p_T range of 34 – 500 GeV. The ratio of the two efficiencies is the finer binned scale factor applied to the $|\eta^\ell|$ measurements.

referred as the nominal control region). Each multijet QCD control region is designed to be enhanced in multijet QCD. To achieve this in the e+jets channel the QCD event selection inverts the isolation requirement in the electron ID given in Tab. 6. For the μ +jets channel the control region event selection requires $0.15 < I_{\text{rel}} < 0.30$. In addition to the isolation requirements both cases also require exactly zero b tagged jets. In these multijet QCD control regions, the contribution of $t\bar{t}$, single top quark, and V+jets background is estimated from simulation to be $\approx 15 - 20\%$. These estimates are subtracted from the data to extract the shape of the multijet QCD in data. The QCD distributions are then scaled to the signal region by a transfer factor derived from simulation

$$t_f = \frac{N_{\text{QCD,MC}}^{\text{Signal}}}{N_{\text{QCD,MC}}^{\text{Control}}}, \quad (6.6)$$

where $N_{\text{QCD,MC}}^{\text{Signal}}$ and $N_{\text{QCD,MC}}^{\text{Control}}$ are the total number of events of the QCD simulation in the signal and control regions respectively. The transfer factor for the control region in each channel is shown in Tab. 9, along with an alternative, orthogonal control region in that channel.

The alternative control region is used to validate the use of the nominal control region. The QCD contribution is estimated in an alternative control region using the shape calculated from the nominal control region. The prediction can then be compared to shape of the data-driven qcd, derived from the alternative control region. The comparison can also yield a sense of the magnitude of the shape and normalisation uncertainties on the data driven QCD estimate. In the e+jets channel the alternative control region uses an electron ID where the conversion veto or missing hits criteria has been inverted and in the μ +jets channel the $0.30 < I_{\text{rel}} < \infty$ region is used. Both still require exactly zero b tagged jets. The data-simulation comparison plots of the event variables in the multijet QCD control regions for the e+jets channel are shown in Figs. 42, 43, 44 and 45 and for the μ +jets channel in Figs. 46, 47, 48 and 49.

Figures 50, 51, 52 and 53 show the data-driven QCD estimates in the alternative control region compared to the QCD calculated by subtracting backgrounds from the data in the alternative control region. The differences between the two predictions show the accuracy of the prediction, which is expected to be similar for the prediction in the signal region. The size of the uncertainty of the QCD prediction in the signal region can therefore be inferred from the discrepancies seen in the alternative control region. The average offset from unity in the ratio of the two QCD yields gives the measure of the normalisation uncertainty, $\approx 60\%$ (gold band) and the variation of the ratio over the range of bins gives a measure of the size of the shape uncertainty, $\approx 30\%$ (grey band). The regions where the ratio is outside these uncertainty bands the expected contribution of QCD is negligible, e.g. at high p_T^{miss} .

In actuality, the multijet QCD uncertainties are propagated, separately for each decay channel, to the final measurement by swapping either the shape or the transfer factor used in the multijet QCD prediction. For the multijet QCD normalisation uncertainty the nominal QCD shape is used with the alternative transfer factor, and for the multijet QCD shape uncertainty, the alternative QCD shape is used with the nominal transfer factor.

Table 9: The transfer factors to the signal region from the four multijet QCD control regions.

Control region	Transfer factor
Non-isolated electron	0.27
Conversion electron	0.03
Non-isolated muon $0.15 < I_{\text{rel}} < 0.30$	0.11
Non-isolated muon $0.30 < I_{\text{rel}} < \infty$	0.08

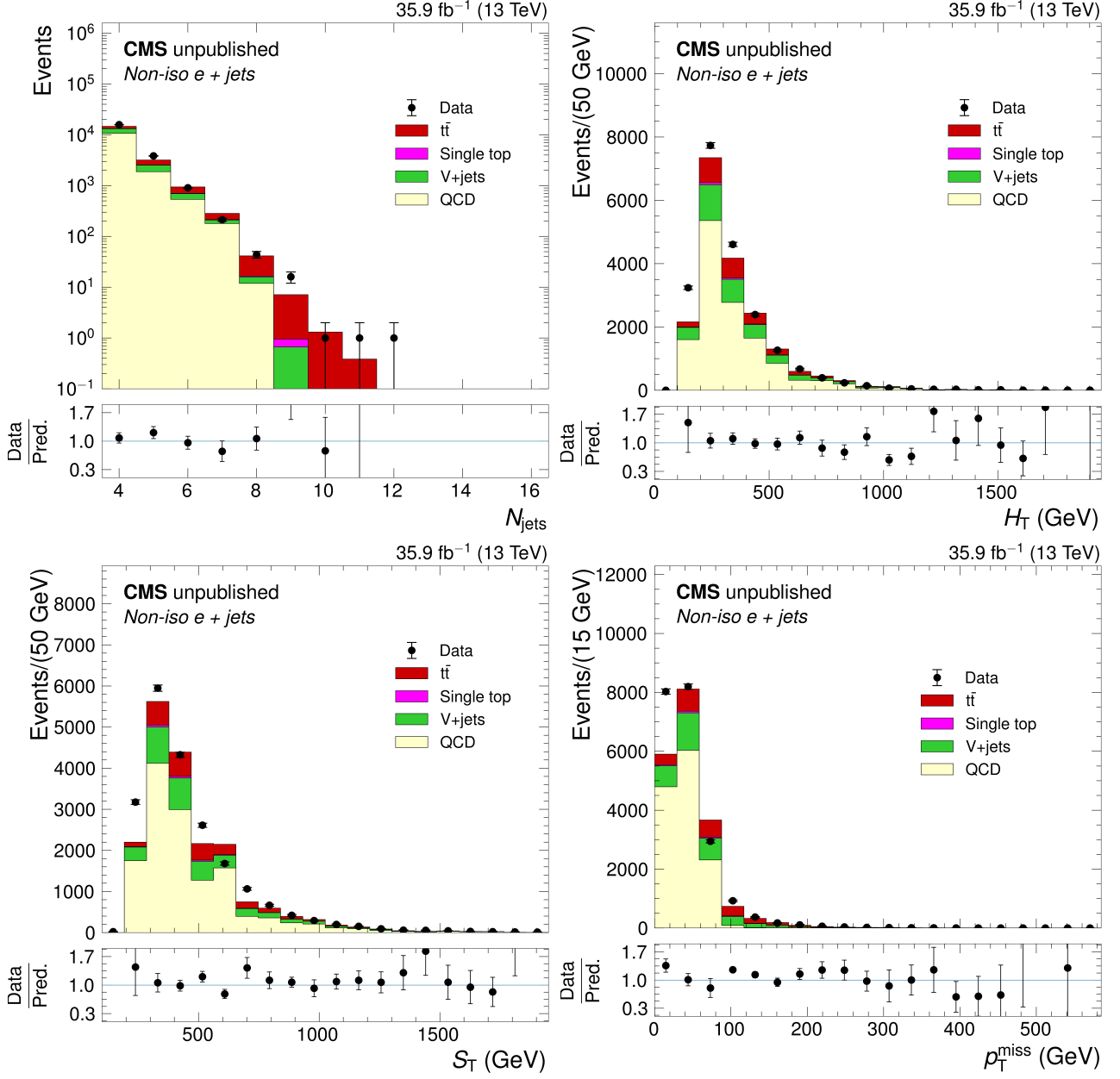


Figure 42: The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the non-isolated electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

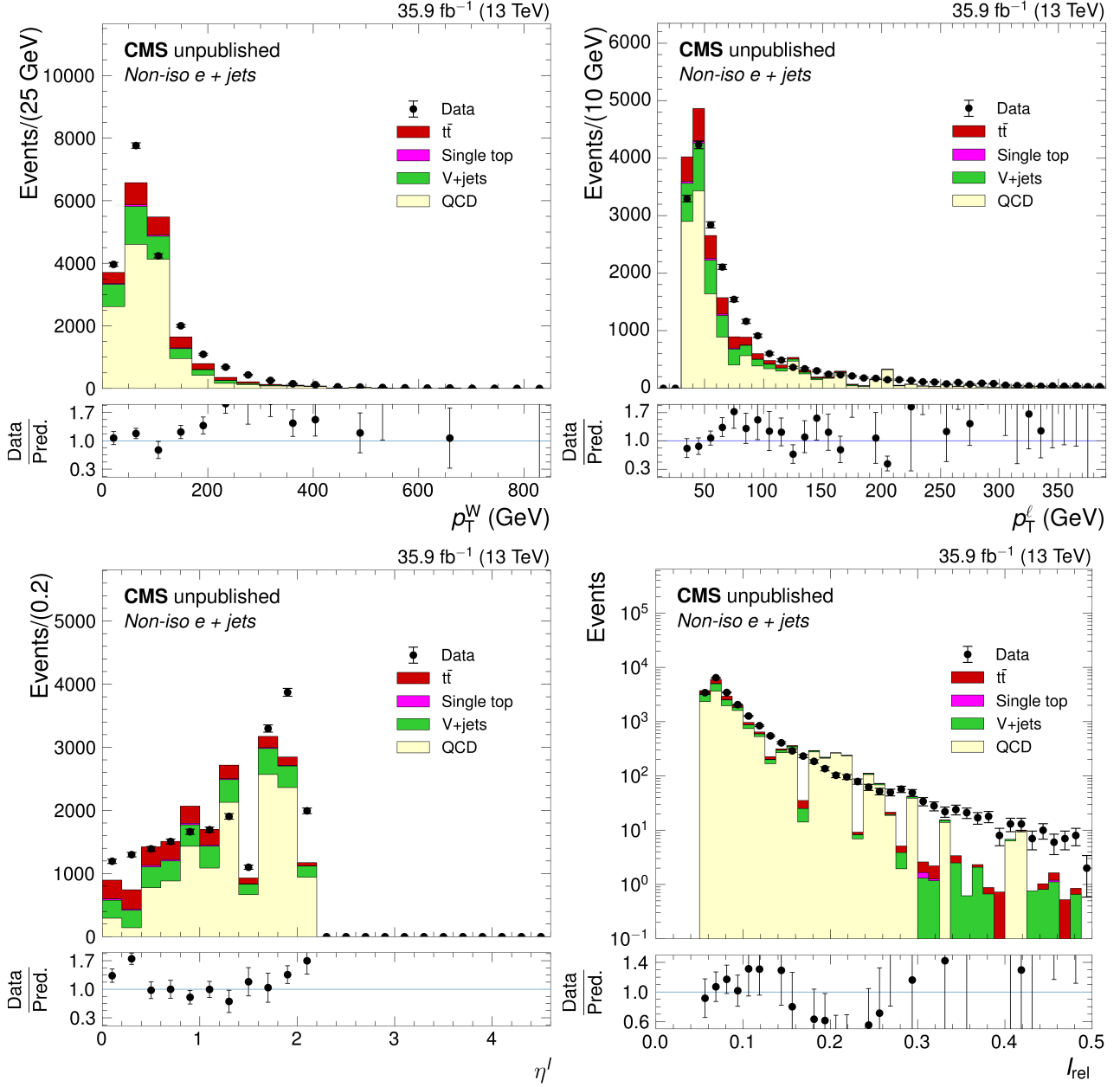


Figure 43: The distributions of the p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables and I_{rel} in the non-isolated electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

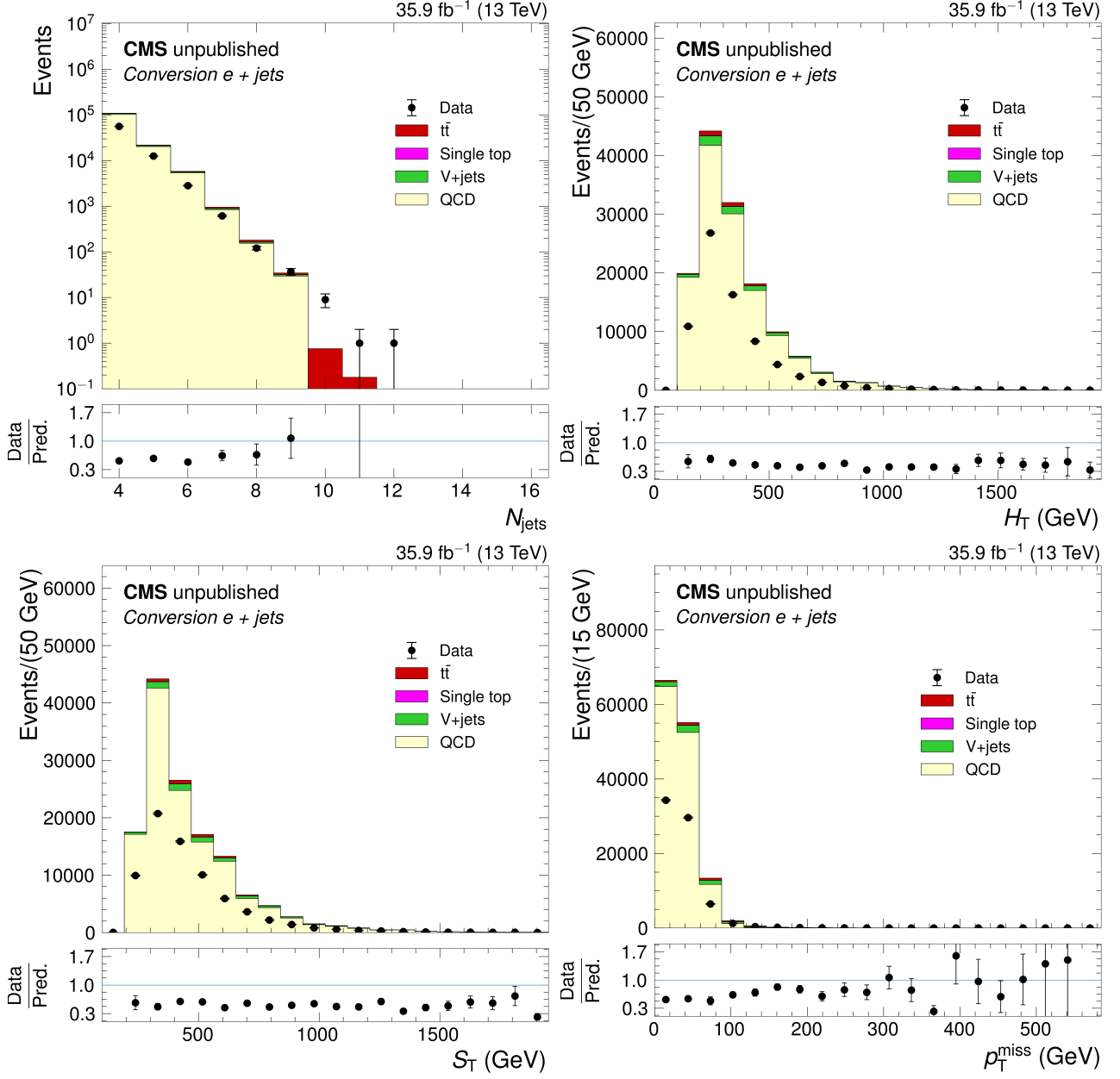


Figure 44: The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the conversion electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

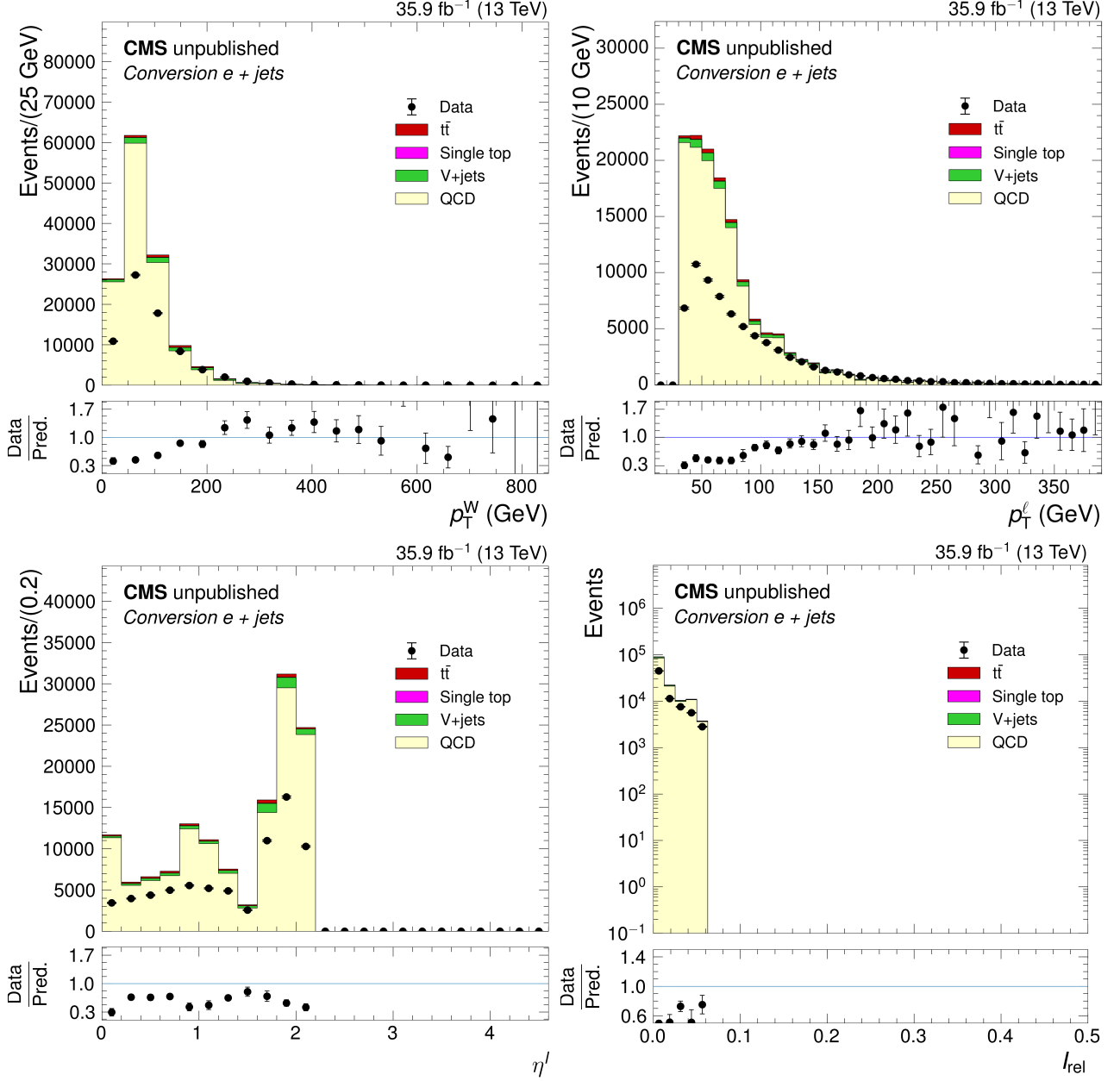


Figure 45: The distributions of the p_T^W , p_T^l and $|\eta^l|$ event variables and I_{rel} in the conversion electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

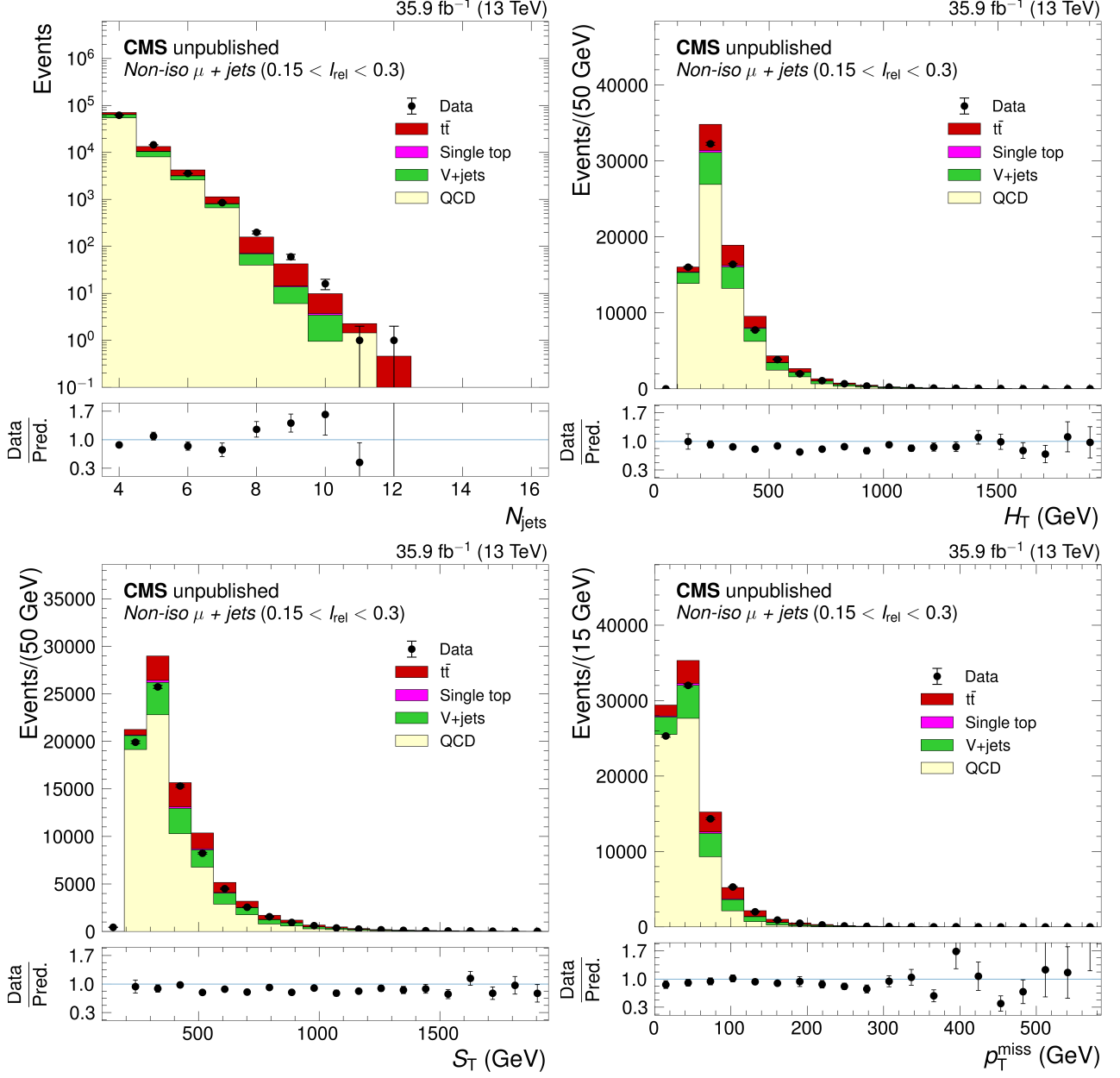


Figure 46: The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables in the non-isolated ($0.15 < I_{\text{rel}} < 0.30$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

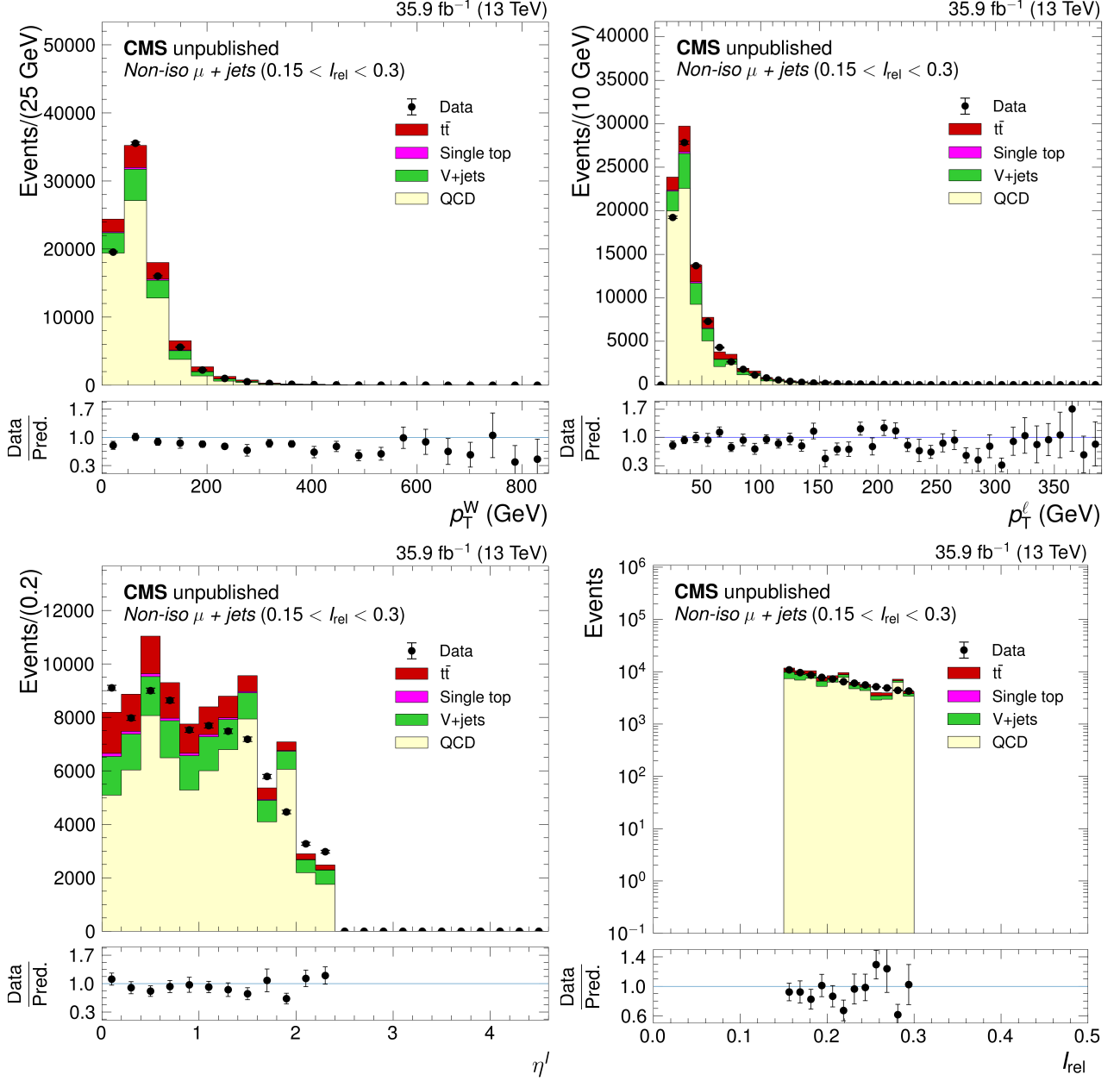


Figure 47: The distributions of the p_T^W , p_T^l and $|\eta^l|$ event variables and I_{rel} in the non-isolated ($0.15 < I_{\text{rel}} < 0.30$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

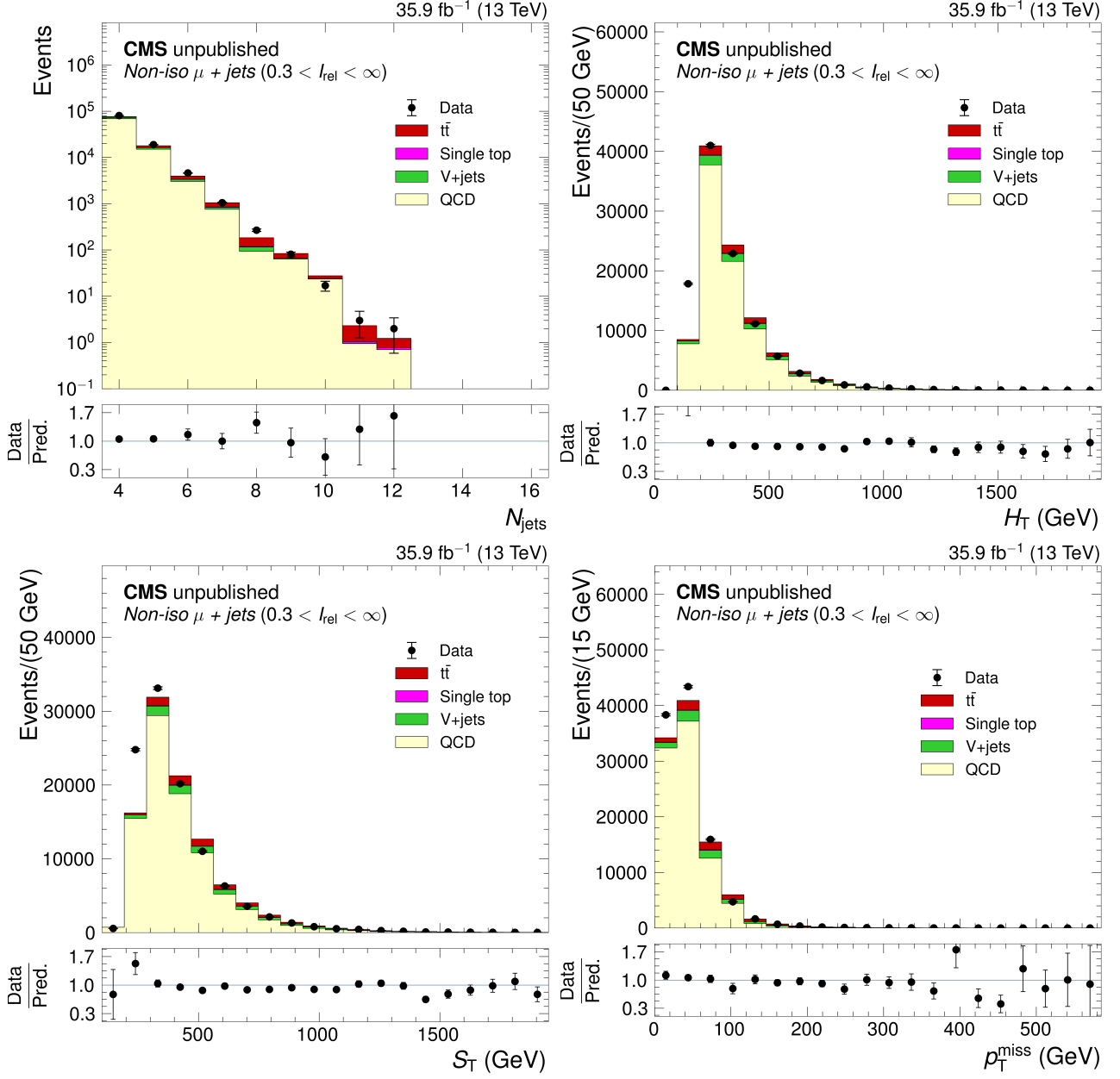


Figure 48: The distributions of the N_{jets} , H_T , S_T and p_T^{miss} event variables and I_{rel} in the non-isolated ($0.30 < I_{\text{rel}} < \infty$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

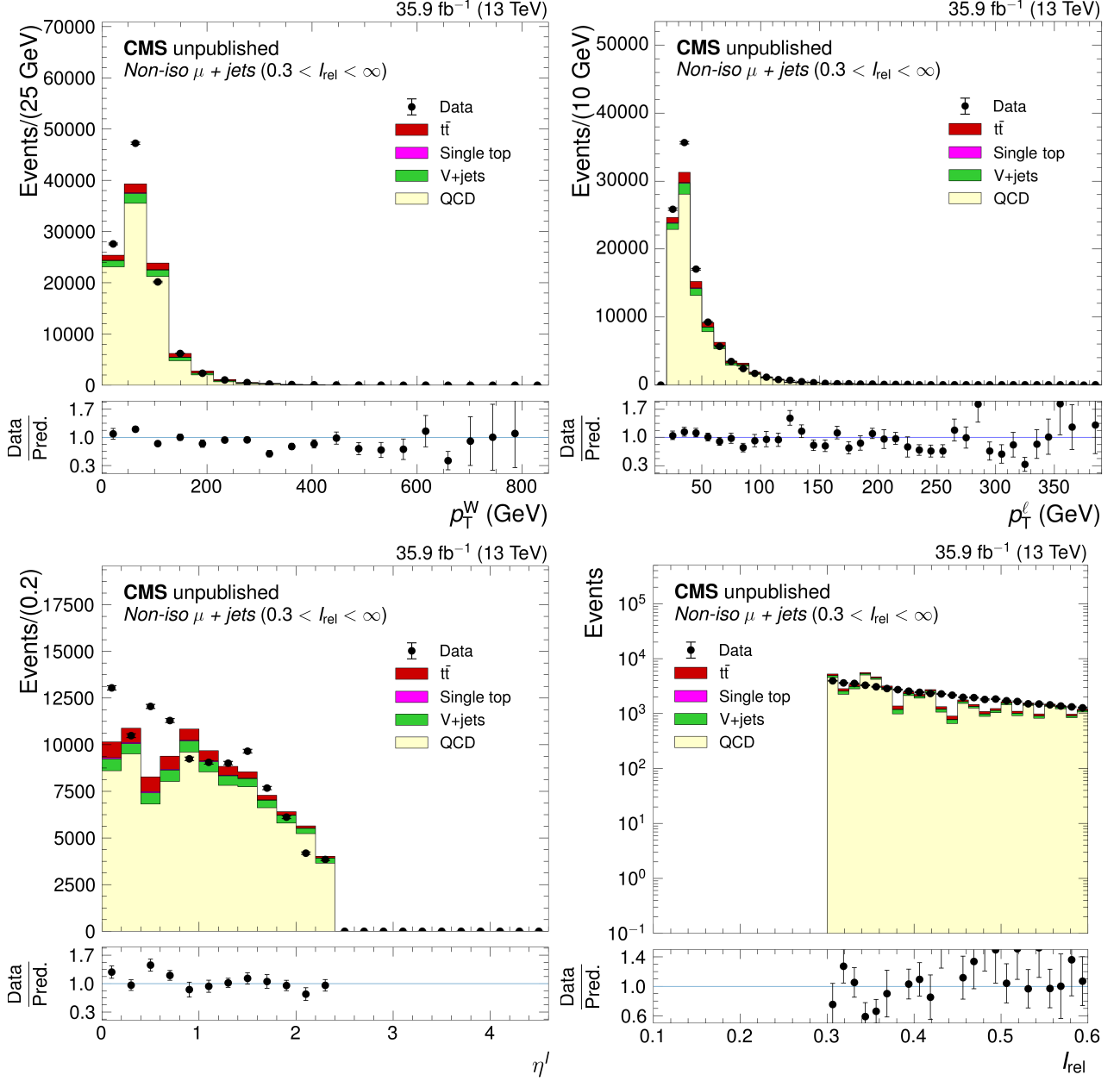


Figure 49: The distributions of the p_T^W , p_T^l and $|\eta^l|$ event variables and I_{rel} in the non-isolated ($0.30 < I_{\text{rel}} < \infty$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

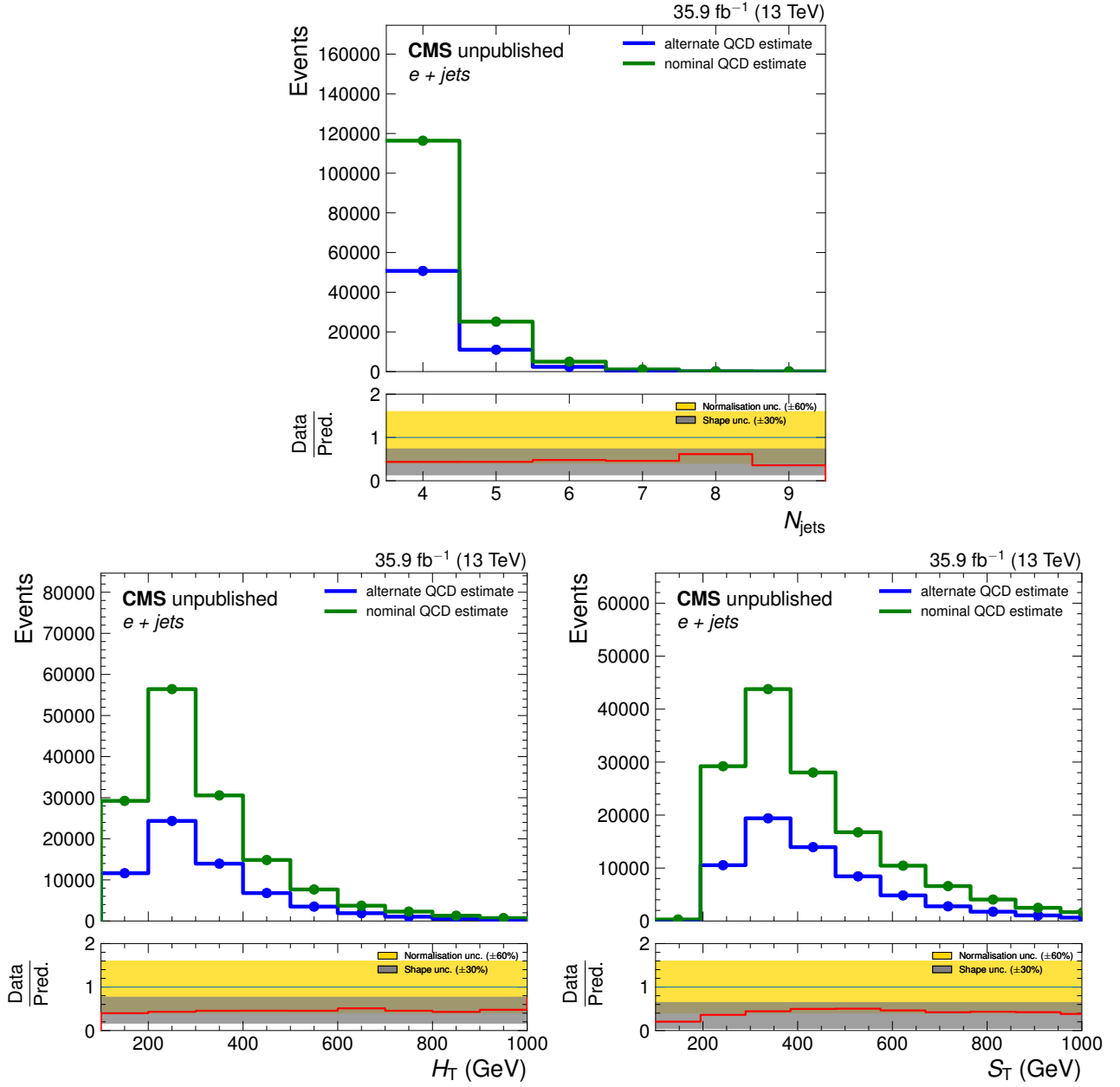


Figure 50: The prediction of multijet QCD for the N_{jets} , H_T and S_T event variables in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be $\pm 60\%$ shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be $\pm 30\%$ shown by the grey band.

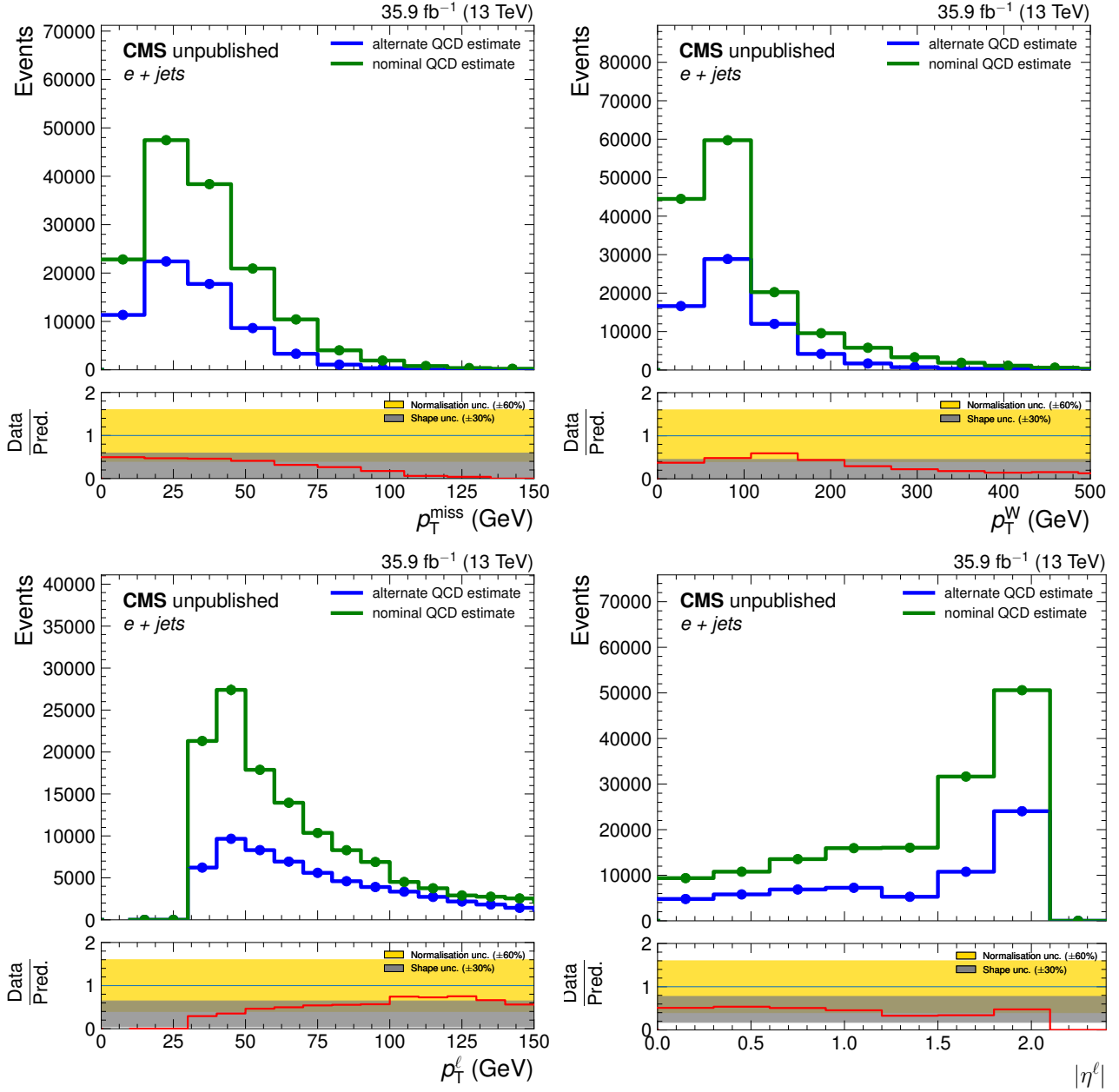


Figure 51: The prediction of multijet QCD for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be $\pm 60\%$ shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be $\pm 30\%$ shown by the grey band.

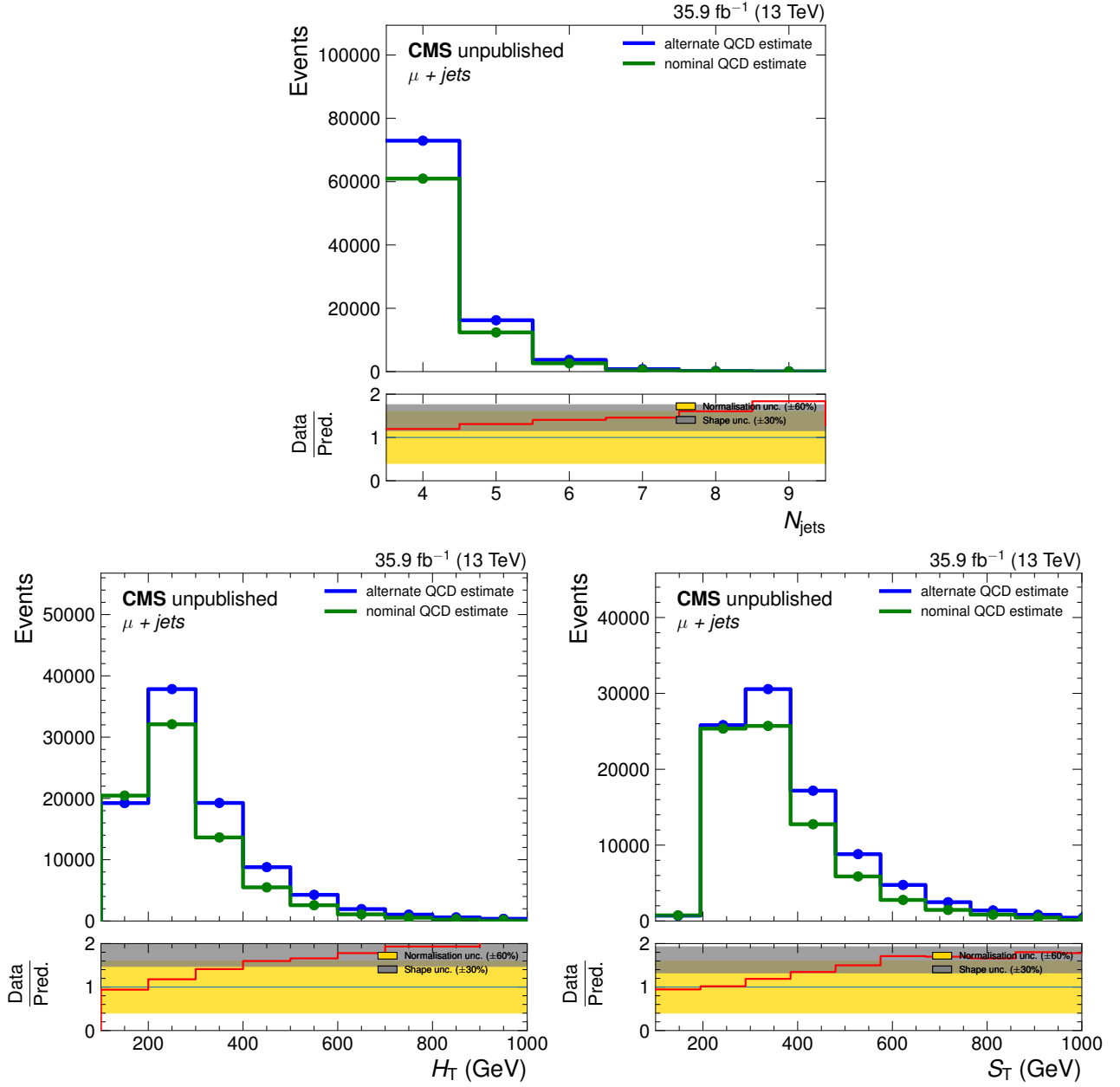


Figure 52: The prediction of multijet QCD for the N_{jets} , H_T and S_T event variables in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be $\pm 60\%$ shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be $\pm 30\%$ shown by the grey band.

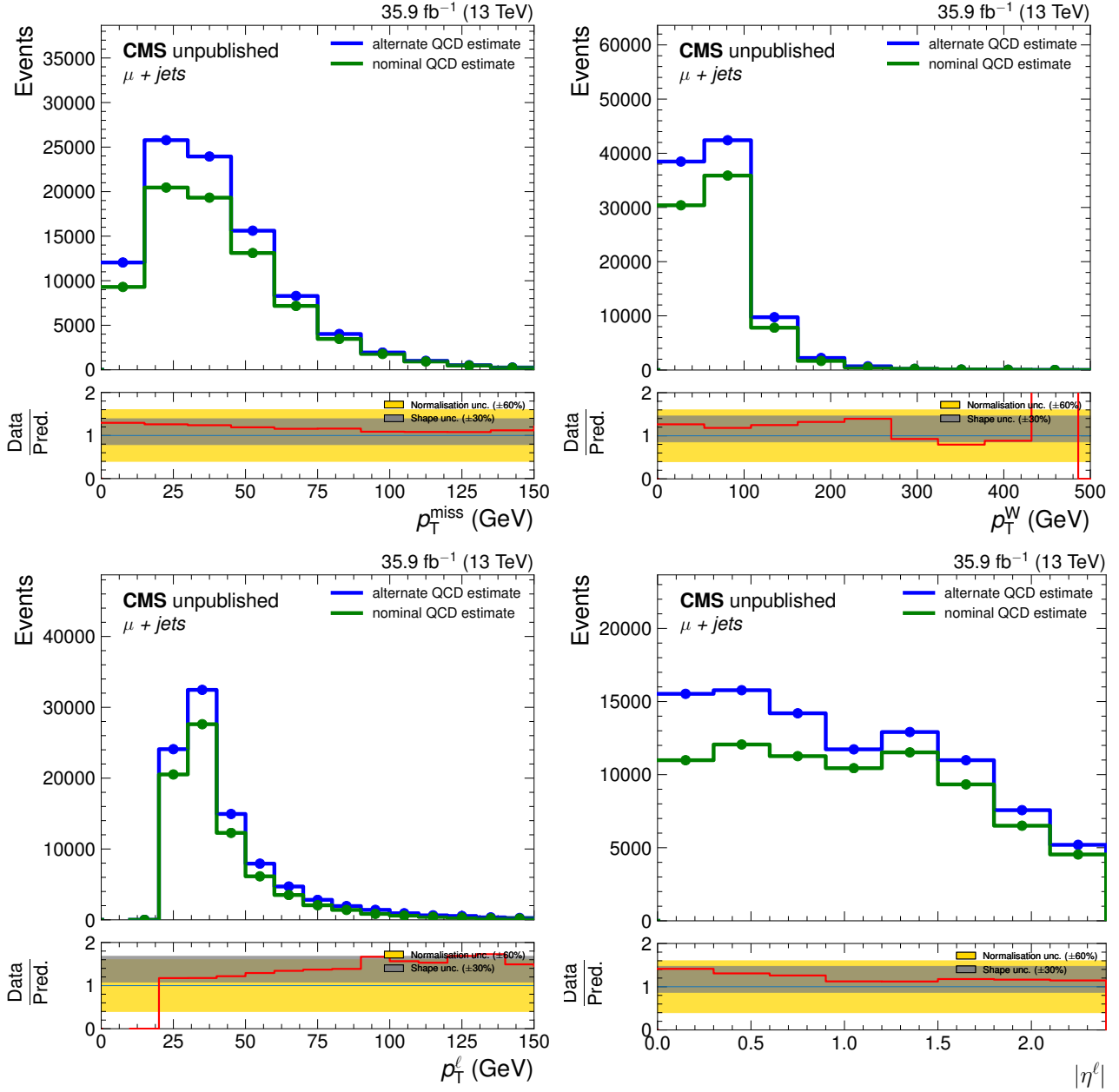


Figure 53: The prediction of multijet QCD for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be $\pm 60\%$ shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be $\pm 30\%$ shown by the grey band.

6.6 Data-simulation agreement in the signal region

Figures 54 and 55 show the agreement between the data and the sum of the signal and background events from simulation in the combined e +jets and μ +jets channels. All corrections previously described have been applied and the multijet QCD contribution estimated from data. A total of 662 381 events are measured in data, compared to 647 892 events in simulation (after averaging the contributions of the data-driven multijet QCD estimates). The simulation and data agree to within 0.2%, indicating that the total cross section is compatible to that stated in Sec. 2.7.1. The simulation predicts the composition of events to be 92.1% $t\bar{t}$ production, 4.4% single top quark production, 2.1% V +jets production and 1.4% from multijet QCD.

The statistical uncertainty in the data is given by vertical error bars. The hatched bands shows the statistical uncertainty in the number of simulated events and experimental uncertainties combined in quadrature. The set of experimental uncertainties include those from the JEC and p_T^{miss} , the luminosity, the pileup, lepton and b tagging scale factors and the background predictions. These uncertainties are explained in more detail in Sec. 8.2.1. The ratios of the event variables that are correlated to the p_T of the top quark show that the p_T^{top} is softer in data than in simulation.

6.6.1 Top quark p_T

The trend seen in variables correlated the p_T^{top} is a well known issue, with many other analyses reporting the same effect, for example in [84, 86, 88, 90]. The trend has been seen in:

- Dilepton, single lepton and hadronic final states
- $\sqrt{s} = 7, 8, 13$ TeV
- Particle, parton and detector level
- 2D differential cross sections

When considering higher-order QCD corrections and/or EWK corrections [165] in the matrix-element, the discrepancy is reduced, albeit not to zero. A correction can be applied when measuring the p_T^{top} distribution, using a reweighting derived from data presented in [88, 166]. The weight is defined as

$$w_{p_T^{\text{top}}} = \sqrt{\text{SF}_{p_T^{\text{top}}} \cdot \text{SF}_{p_T^{\text{antitop}}}}, \quad (6.7)$$

where the scale factor is measured to be

$$\text{SF}_{p_T} = e^{0.0615 - 0.0005 \cdot p_T}. \quad (6.8)$$

If the reweighting is applied to this measurement, the slope observed in the ratio of these variables is indeed corrected, as shown in Fig. 56.

However, the choice of applying the reweighting of the p_T^{top} has consequences. It can have a significant impact on other variables, especially the mass of the $t\bar{t}$ system, $M_{t\bar{t}}$. As the measurements presented in this thesis do not measure the p_T^{top} distribution, the correction is not applied. It is, however, propagated to the final measurements as a systematic uncertainty, and, looking ahead to Figs 77, 78, 79 and 80, the effect of the p_T^{top} reweighting applied to the POWHEG+PYTHIA model can be compared to the differential cross section measurements.

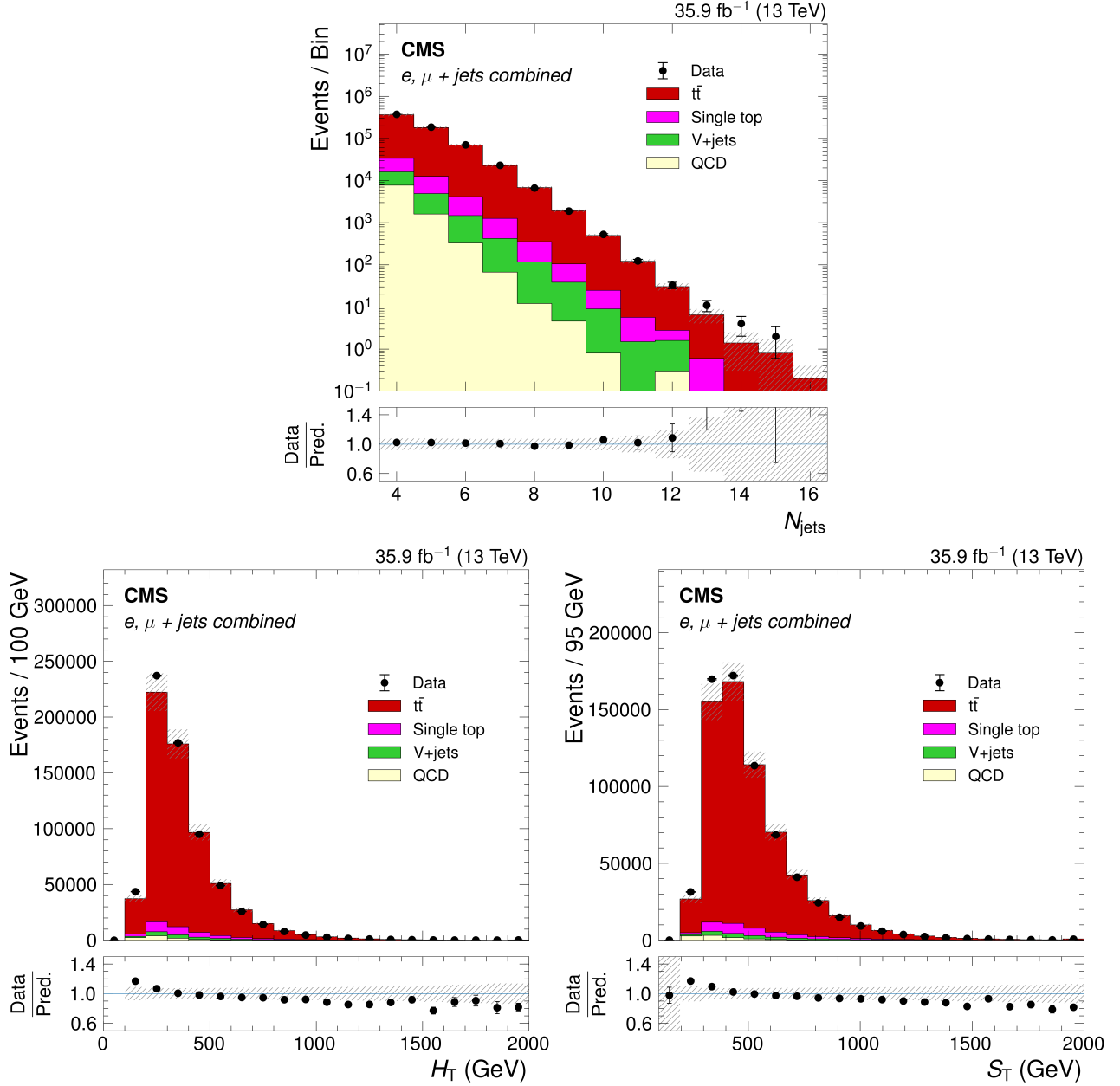


Figure 54: The distributions of N_{jets} , H_T and S_T after full event selection. The $t\bar{t}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band.

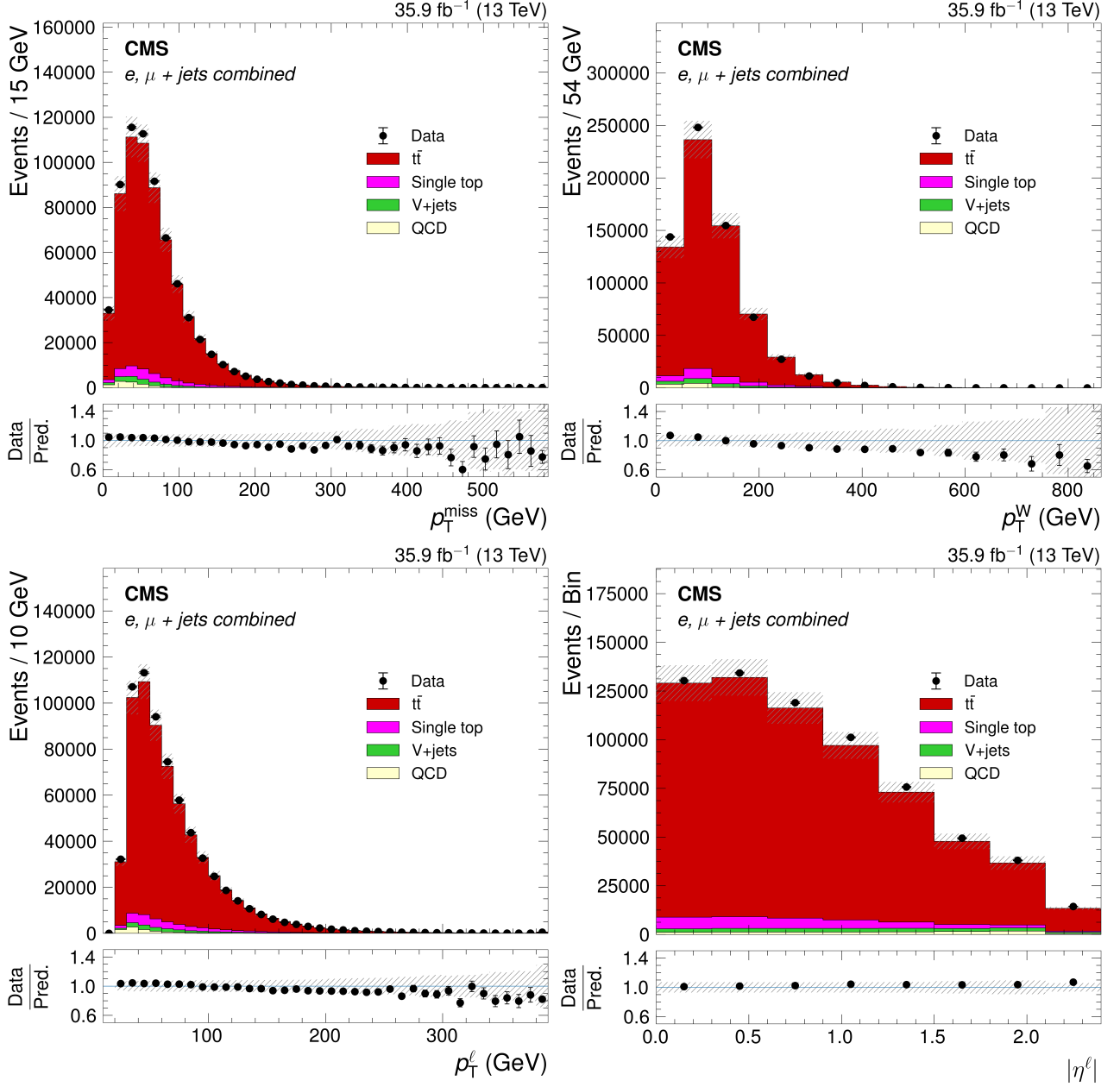


Figure 55: The distributions of p_T^{miss} , p_T^W , p_T^l and $|\eta^l|$ after full event selection. The $t\bar{t}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band.

6.6.2 Calculating the yield of top quark pairs

The $t\bar{t}$ yield can be extracted from data by subtracting off the simulated backgrounds and data-driven multijet QCD estimate. Events that are reconstructed but do not enter the visible phase space at particle level, detailed in Sec. 7.1, are also subtracted. The extracted $t\bar{t}$ yields are shown per bin for each event variable, in both the e+jets and μ +jets channels, in Tab. 10. The difference in the total $t\bar{t}$ yield between the event variables originates from the differences in the predictions of the QCD background for each variable.

The yields are still dependent on the resolution, acceptance and efficiency of the detector and would make any cross section measurement comparison only valid where both include the modelling of the CMS experiment. This is time consuming in terms of simulation and does not allow like-for-like comparisons with other experiments. The detector effects are corrected for in the process known as unfolding.

Table 10: The yield of $t\bar{t}$ estimated from the data in each bin for each event variable, after the backgrounds have been subtracted and the events not entering the phase space removed. The combined total is also shown, where the differences are due to the multijet QCD estimation.

	N_{jets}		H_T		S_T		p_T^{miss}		p_T^W		p_T^ℓ		$ \eta^\ell $	
	e+jets	μ +jets	e+jets	μ +jets	e+jets	μ +jets	e+jets	μ +jets	e+jets	μ +jets	e+jets	μ +jets	e+jets	μ +jets
Bin 1	119093	199104	24766	43648	17251	34976	93510	148677	36616	73313	24136	95549	46964	68369
Bin 2	62708	102831	43925	76303	46290	82681	90590	154097	81633	141297	56492	86796	47170	71588
Bin 3	24527	39321	46982	78608	50965	85795	26419	45211	60153	91306	44433	61162	42412	62465
Bin 4	8032	12904	36136	59113	38548	60701	5383	8450	27720	38543	31077	40540	35970	52702
Bin 5	2392	3668	24622	38966	26213	39828	1292	2050	8535	11006	20509	26049	22080	43279
Bin 6	975	1392	16473	25672	15801	23594	588	825	2280	2818	13232	16857	12358	29249
Bin 7	—	—	9941	15164	9751	13932	—	—	870	1040	9155	10745	7988	14065
Bin 8	—	—	6159	9113	5671	7876	—	—	—	—	5938	6966	3009	17546
Bin 9	—	—	3700	5401	3115	4335	—	—	—	—	3889	4565	—	—
Bin 10	—	—	2151	3100	1863	2453	—	—	—	—	2653	3013	—	—
Bin 11	—	—	1107	1670	998	1327	—	—	—	—	1899	2100	—	—
Bin 12	—	—	715	887	589	742	—	—	—	—	1228	1398	—	—
Bin 13	—	—	861	1222	565	755	—	—	—	—	902	988	—	—
Bin 14	—	—	—	—	—	—	—	—	—	—	585	696	—	—
Bin 15	—	—	—	—	—	—	—	—	—	—	595	614	—	—
Bin 16	—	—	—	—	—	—	—	—	—	—	494	559	—	—
Bin 17	—	—	—	—	—	—	—	—	—	—	538	642	—	—
Total (combined)	596084		596123		596128		596137		596120		596135		593396	

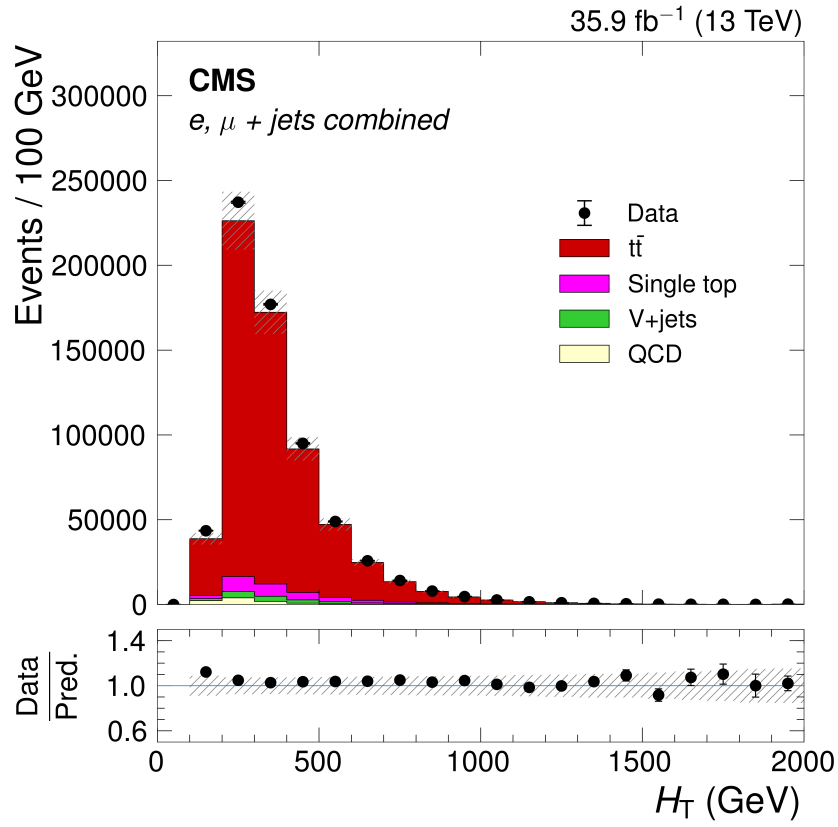


Figure 56: The distribution of the H_T variable after full event selection and the application of top quark p_T reweighting. With the application of the reweighting the ratio is now flat.

CHAPTER 7

Removing the CMS thumbprint

The detectors of the CMS experiment have a finite resolution, which can lead to mis-measurements of the kinematic properties of particles and therefore bin-to-bin migrations when reconstructed. Additional effects originating from the acceptance of the detector and the efficiencies of the sensitive components of the detector are also introduced. Together, these effects are known as the response of the detector.

When comparing to measurements from other experiments or to simulations without the modelling of the detector, these effects, derived in simulation, need to be removed. This process is known as unfolding. There are two common approaches to the level of unfolding. Firstly, one can unfold to parton level which extrapolates the results with respect to the constituent partons. This has large theoretical uncertainties stemming from the modelling of the parton shower. In addition, these measurements are usually also extrapolated to the full kinematic phase space. The second option is to present the results to particle level, i.e. with respect to the stable particles produced after the shower modelling, where stable particles are, for this purpose, defined as particles with a lifetime longer than 30 ps, in a phase space close to that accessible by the CMS detector.

7.1 Particle level measurement

This thesis presents measurements at particle level in a phase space chosen to closely resemble that used to select events in data. Particle-level objects are used to define the phase space which is identical for both the e +jets and μ +jets channels which allows for a consistent combination after unfolding. Particle-level objects used to define the phase space are constructed from stable particles produced in simulation by the event generator but before the detector interactions are modelled.

The generator-level descriptions of the particles are based on the Robust Independent Validation of Experiment and Theory (RIVET) framework [167], following prescriptions given in [168]. Simulated electrons and muons not originating from a hadron or a quark are used to define electrons and muons at particle level. Photons close to the lepton are assumed to have radiated from the lepton, and are clustered with it using the anti- k_T algorithm with a cone radius, $R = 0.1$. Particle-level jets are constructed by clustering all stable particles, excluding those used in defining the leptons, with the anti- k_T algorithm with $R = 0.4$. As the b

hadronises into a B hadron and travels up to a few cm through the detector before decaying, a particle-level jet is tagged as originating from a b quark if the initiating particle is a B hadron. The particle-level p_T^{miss} is calculated from all stable visible particles.

The common, visible phase space to be used in the combination of the e+jets and μ +jets decay channels is given by a single particle-level electron or muon with $p_T > 26$ GeV and $|\eta| < 2.4$. Events with additional leptons of $p_T > 15$ GeV and $|\eta| < 2.4$ are not permitted. The particle-level jet selection differs slightly from that in data. The three leading particle-level jets must have a $p_T > 30$ GeV, however the p_T requirement for the fourth jet is relaxed to > 20 GeV. Two of the particle-level jets need to be tagged as originating from a b quark. The variables H_T , S_T and N_{jets} are subsequently calculated with respect to all jets with $p_T > 20$ GeV.

As stated in Sec. 6.6.2, the yield of $t\bar{t}$ events for each bin in data is obtained by subtracting the contribution of each background process. In addition, the contribution of $t\bar{t}$ events which satisfy the selection criteria, but do not enter the visible phase space at particle level, is estimated from simulation and subtracted from the data. This accounts for $\sim 7\%$ of all $t\bar{t}$ events and are predominately those in which one of the jets fails the particle-level jet selection, but passes the reconstructed jet selection because of the resolution of the detector. The relaxed particle-level jet selection reduced this fraction from $\sim 20\%$ to create the largest possible data sample. The effect on the total uncertainty from this additional extrapolation is negligible. No selection is applied on the decay channel of the top quarks, so the phase space does not exclusively contain single electron or muon $t\bar{t}$ events. In particular, there are contributions from events where one top quark decays to a tau lepton and subsequently to an electron or muon, or where both top quarks decay leptonically but one lepton is not within the particle-level acceptance.

7.2 Choice of bins

Events can migrate between bins of the measurement because of the finite resolution of the CMS detector, i.e. the reconstructed variable can have a different value, and so can be in a different bin, compared to the particle-level variable. The choice of binning used in the migration matrix between the reconstructed distributions and the particle-level distributions can be used to minimise the migrations. The binning scheme for each of the event variables, with the exception of the jet multiplicity whose binning is naturally defined, is created by iteratively adding together fine bins of that variable until a set of binning criteria is met.

The primary criteria to reduce the level of migration can be represented by the purity p and stability s of the bin which are defined as

$$p^i = \frac{N_{\text{rec\&gen}}^i}{N_{\text{rec}}^i} \quad (7.1)$$

and

$$s^i = \frac{N_{\text{rec\&gen}}^i}{N_{\text{gen}}^i}, \quad (7.2)$$

where $N_{\text{rec\&gen}}^i$ is the number of events generated and reconstructed in bin i and N_{rec}^i (N_{gen}^i) is the number of events reconstructed (generated) in bin i . Purity measures the effect of migrations into bins and stability migrations out of bins. By requiring p^i and $s^i > 0.6$ an acceptable balance between the minimisation of the migration of events between bins and the retention of information is obtained. The purity and stability selections are reduced to 0.5 for p_T^{miss} , due to its naturally low resolution. Figures 57 and 58 show the purity and stability of bins in the e+jets channel. The corresponding purities and stabilities of the bins in the μ +jets channel are shown in Figs. 90 and 91 in App. A.

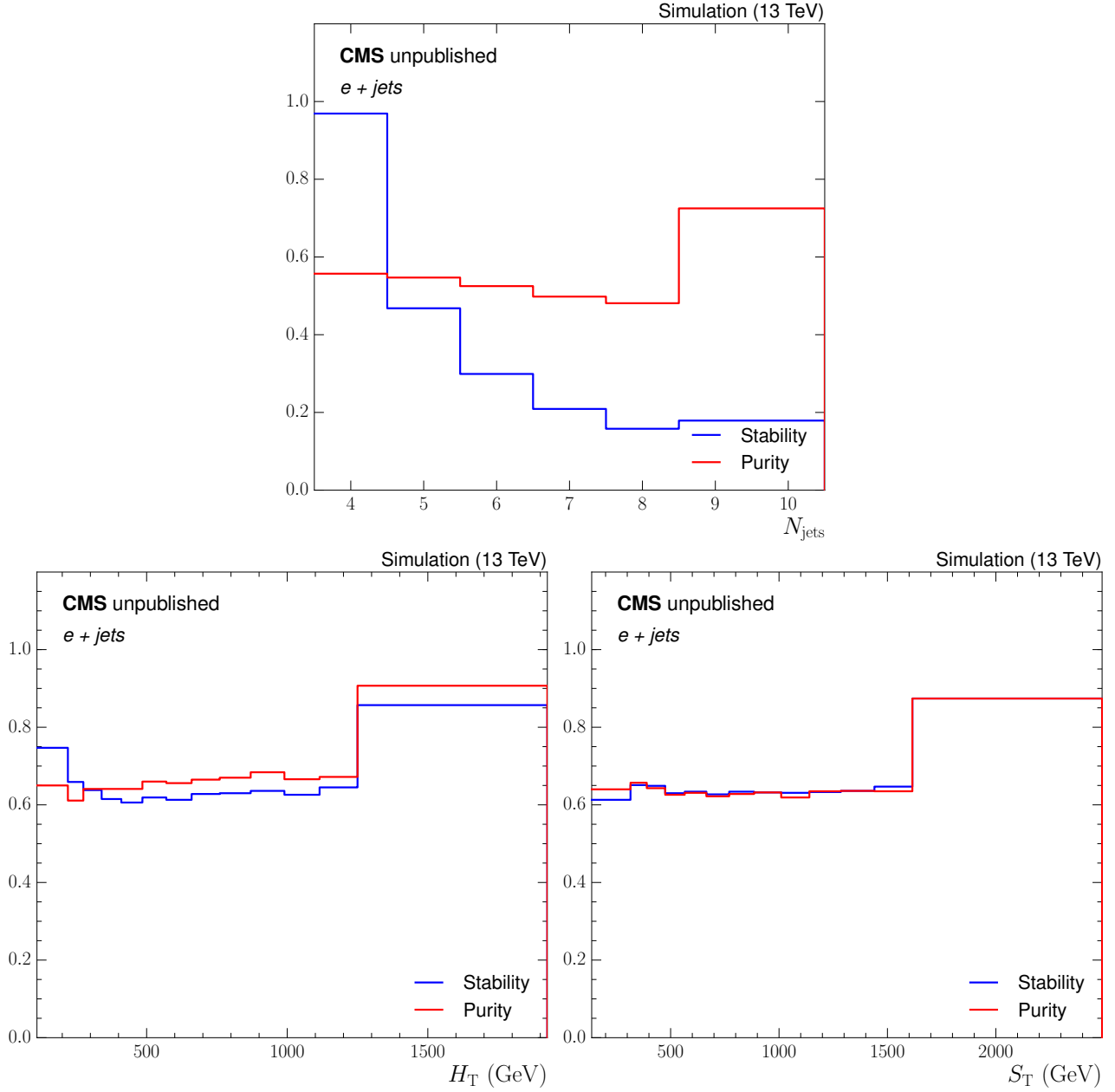


Figure 57: The purity and stability of the bins in the $e+jets$ channel, measured using the POWHEG+PYTHIA simulation sample, for the N_{jets} , H_T and S_T event variables.

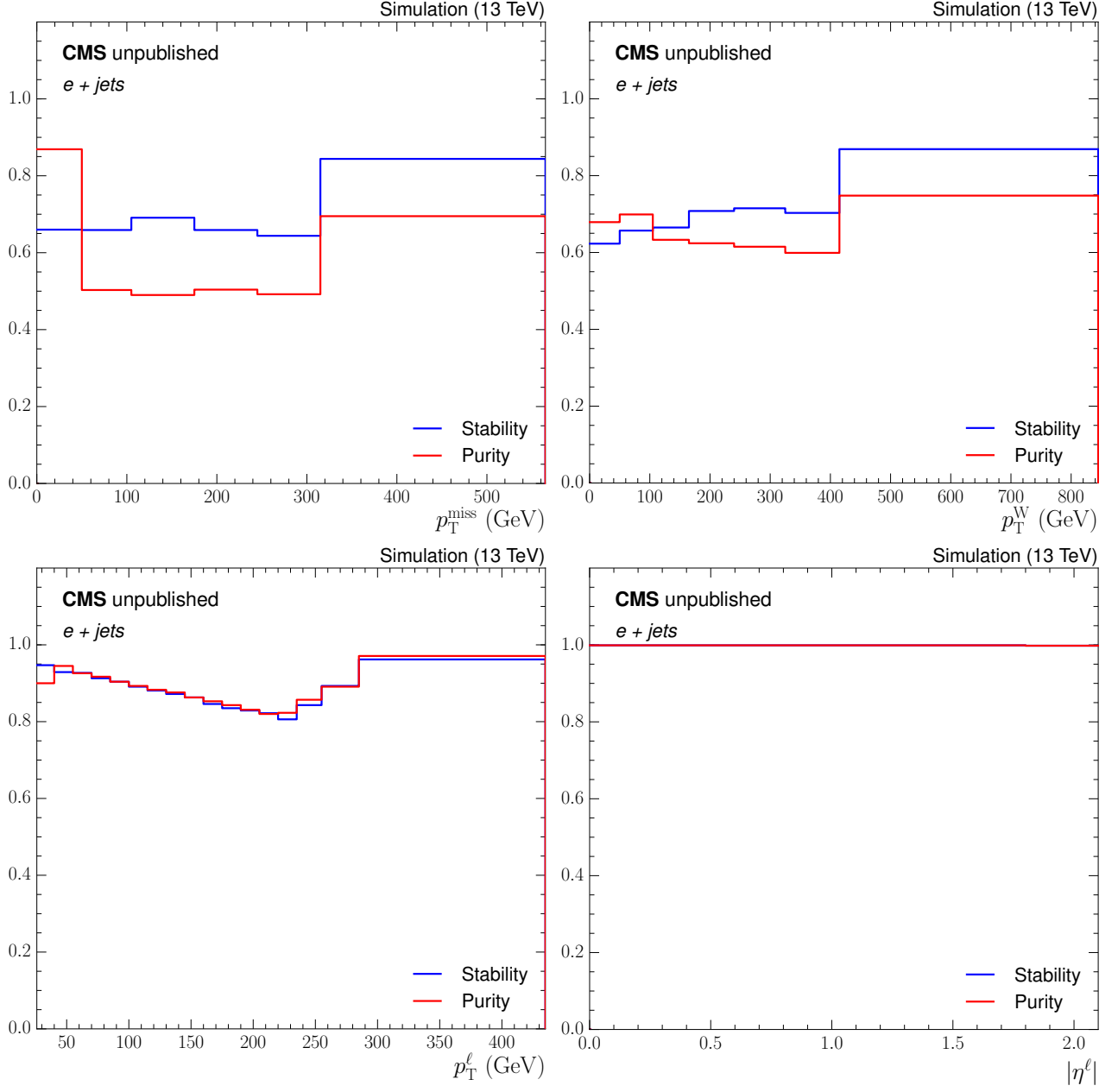


Figure 58: The purity and stability of the bins in the e+jets channel, measured using the POWHEG+PYTHIA simulation sample, for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables.

On top of purity and stability, a few other restrictions are required for the choice of binning. Each bin of a variable must be at least as wide as the typical resolution above and below the centre of that bin. The resolution is defined as the standard deviation of the Gaussian distributed residuals ($V_{\text{Reco}} - V_{\text{Gen}}$) of a variable that have been calculated from the POWHEG+PYTHIA simulation. Figure 59 shows an example of a comparison between the resolution, half-bin width and the absolute values of the residual. It uses events from both the e +jets and μ +jets channels, in the H_T bin spanning $220 - 275$ GeV. Tables 11 and 12 show the resolution and bin width

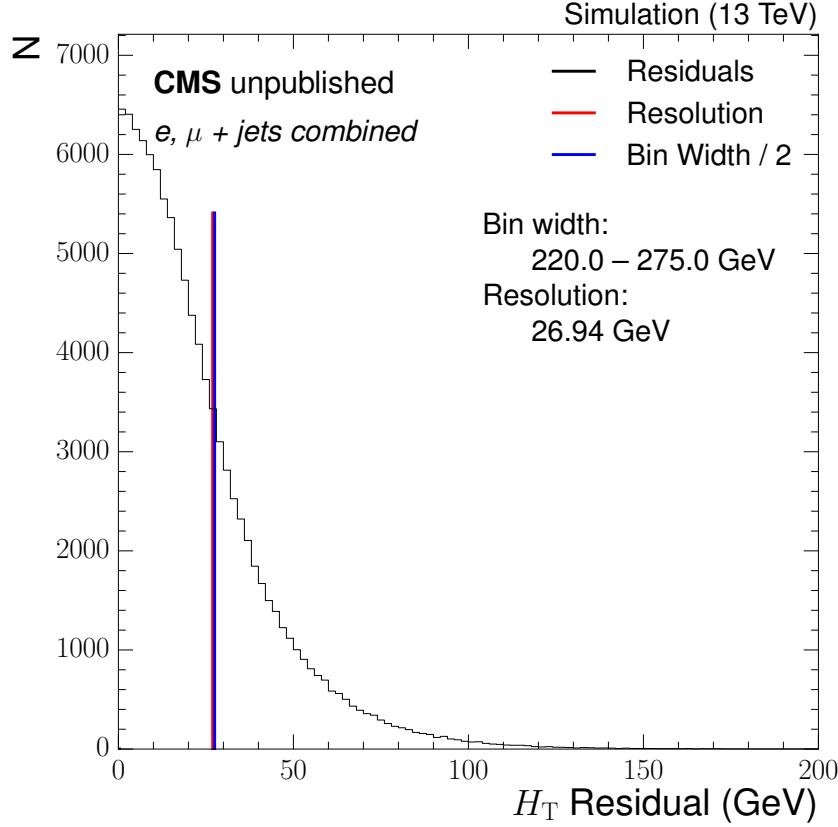


Figure 59: The residual distribution for the H_T event variable using events from both the e +jets and μ +jets channels. The resolution is given as the vertical red line and compared to the half-bin width, given as the vertical blue line.

comparisons for each bin of each variable in the e +jets and μ +jets channels respectively and the complete set of residual distributions are shown in App. B. The number of reconstructed $t\bar{t}$ events in each bin, as estimated from simulation, needs to be at least 500, which corresponds to a maximum statistical uncertainty of $\approx 5\%$. If the final bin contains less than 500 events it is merged with the previous bin. This requirement is to ensure the migration matrices used in the unfolding procedure do not introduce a bias based on the statistical fluctuations of the simulation.

Figure 60 shows an example of the fine bin migration matrix between the reconstructed distribution and the particle-level distribution for the H_T event variable in the e +jets channel on the left panel and μ +jets channel on the right panel. The best binning scheme calculated is overlaid in red. The migration matrices for the e +jets channel, calculated using the common binning scheme, are shown in Figs. 61 and 62. Similarly, the migration matrices for the μ +jets channel are shown in Figs. 101 and 102 or App. C. The binning requirements result in close-to-diagonal migration matrices with an acceptable statistical uncertainty.

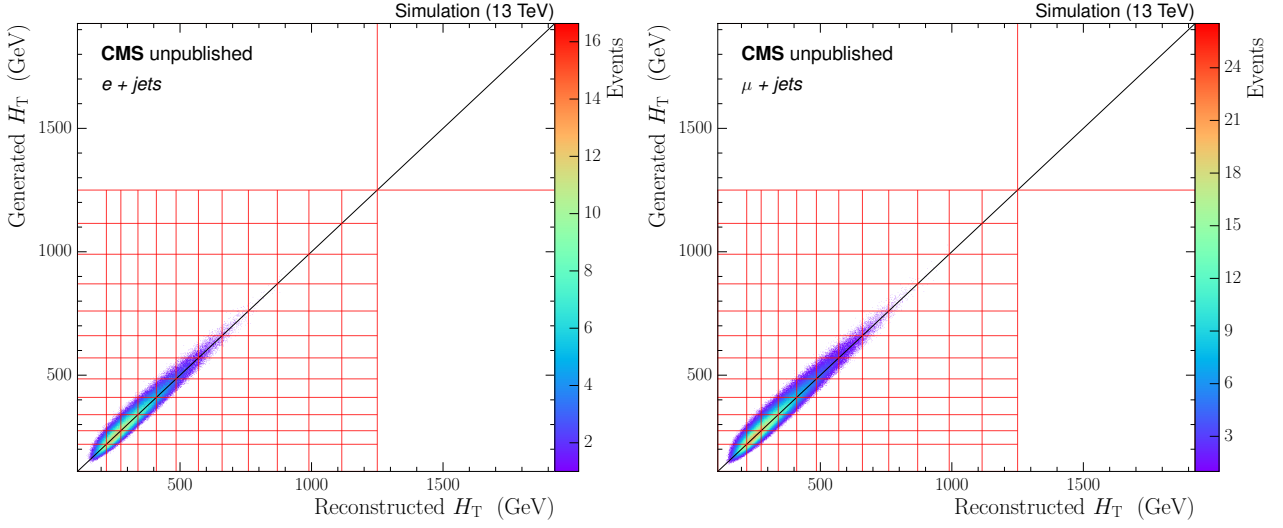


Figure 60: The left panel shows the mapping between the reconstructed and particle-level distributions of the H_T event variable in the e +jets channel and the right panel in the μ +jets channel. The chosen common binning scheme passing all criteria is overlaid in red.

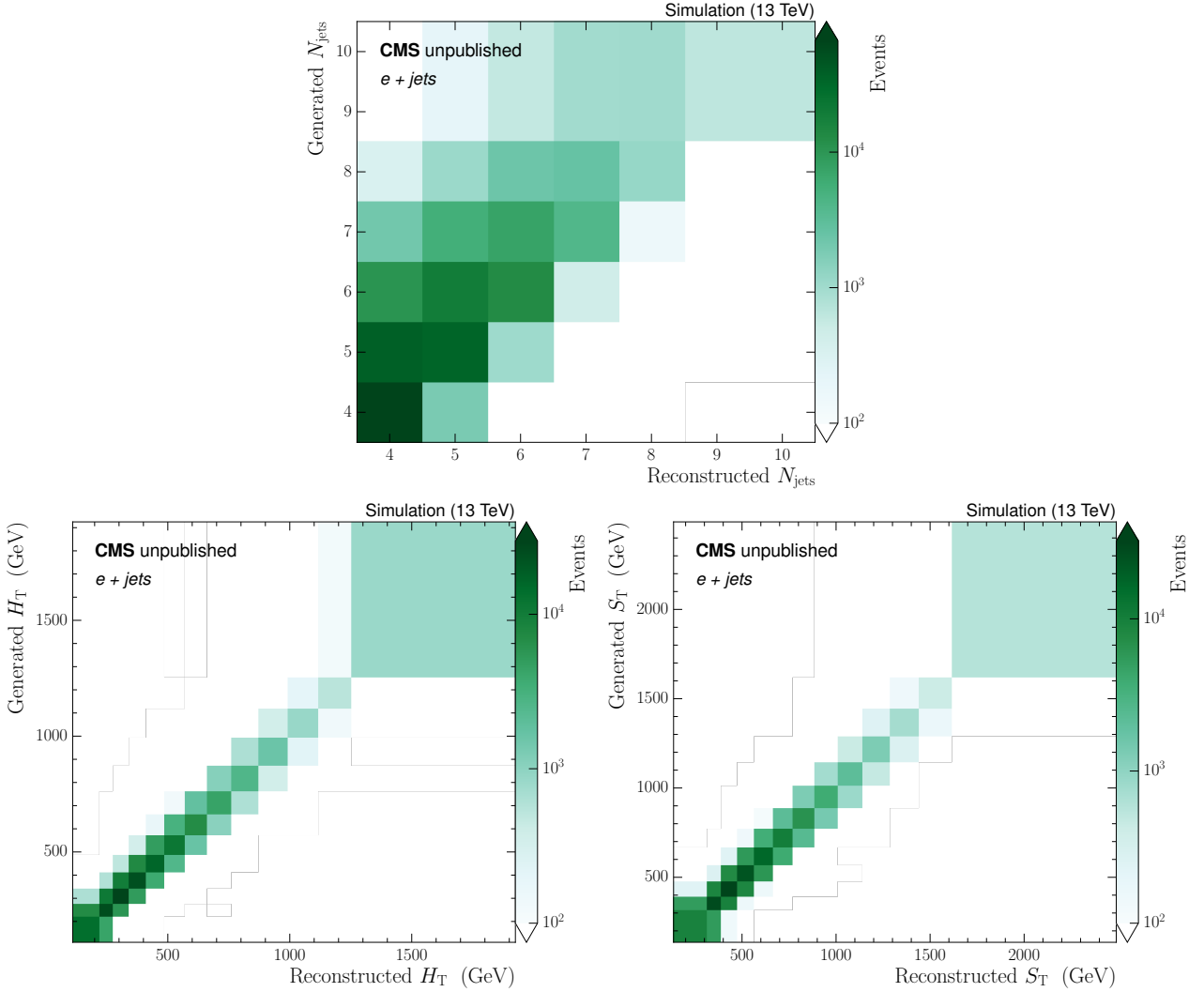


Figure 61: The set of migration matrices calculated using the POWHEG+PYTHIA simulation sample for the N_{jets} , H_T and S_T event variables in the e +jets channel using the common binning scheme.

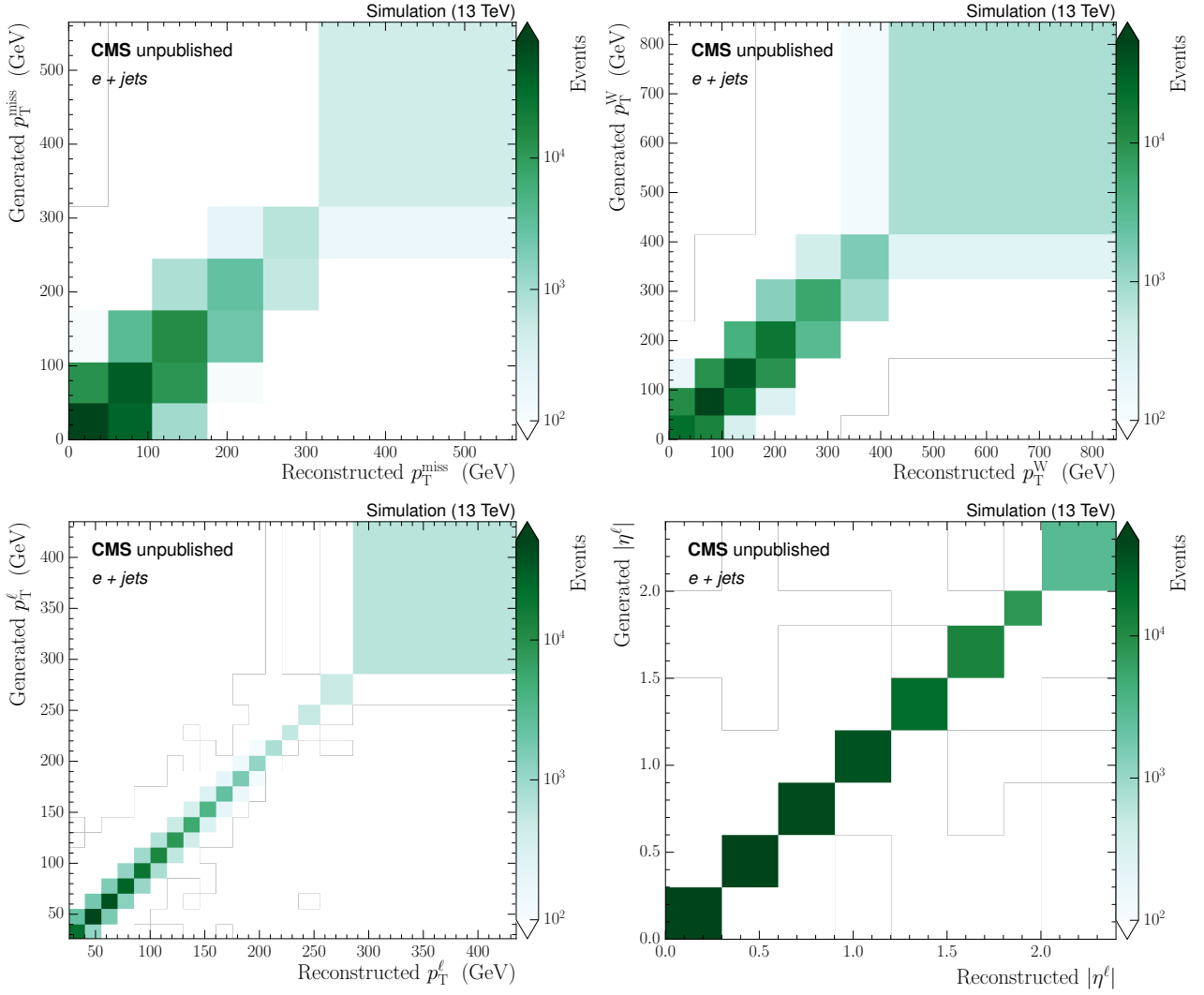


Figure 62: The set of migration matrices calculated using the POWHEG+PYTHIA simulation sample for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the $e + \text{jets}$ channel using the common binning scheme.

Table 11: A comparison of the bin width and the resolution in each bin for all event variables in the electron channel.

N_{jets}		H_T		S_T		p_T^{miss}		p_T^W		p_T^{ℓ}		$ p_T^{\ell} $	
Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution
1.0	1.0	110.0	26.4	179.0	33.4	50.0	17.1	50.0	21.0	14.0	0.9	0.30	< 0.1
1.0	1.0	55.0	27.1	75.0	34.2	55.0	28.5	55.0	22.6	15.0	1.1	0.30	< 0.1
1.0	1.0	65.0	29.4	85.0	38.4	70.0	32.0	60.0	26.4	15.0	1.3	0.30	< 0.1
1.0	1.0	70.0	31.3	90.0	41.9	75.0	32.5	75.0	30.2	15.0	1.4	0.30	< 0.1
1.0	1.0	75.0	33.8	100.0	45.2	70.0	33.8	85.0	33.6	15.0	1.6	0.30	< 0.1
2.0	1.0	85.0	36.0	105.0	48.8	250.0	34.3	90.0	36.6	15.0	1.8	0.30	< 0.1
—	—	90.0	38.5	115.0	52.1	—	—	430.0	39.7	15.0	1.9	0.20	< 0.1
—	—	100.0	41.1	125.0	56.2	—	—	—	—	15.0	2.1	0.40	< 0.1
—	—	110.0	43.7	130.0	60.0	—	—	—	—	15.0	2.4	—	—
—	—	120.0	47.0	145.0	63.6	—	—	—	—	15.0	2.5	—	—
—	—	125.0	50.0	155.0	68.9	—	—	—	—	15.0	2.7	—	—
—	—	135.0	52.8	175.0	73.2	—	—	—	—	15.0	2.9	—	—
—	—	675.0	58.0	875.0	80.7	—	< 0.1	—	—	15.0	3.1	—	—
—	—	—	—	—	—	—	—	—	—	15.0	3.3	—	—
—	—	—	—	—	—	—	—	—	—	20.0	3.5	—	—
—	—	—	—	—	—	—	—	—	—	30.0	3.8	—	—
—	—	—	—	—	—	—	—	—	—	150.0	4.6	—	< 0.1

Table 12: A comparison of the bin width and the resolution in each bin for all event variables in the muon channel.

N_{jets}		H_T		S_T		p_T^{miss}		p_T^W		p_T^{ℓ}		$ p_T^{\ell} $	
Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution	Bin width	Resolution
1.0	1.0	110.0	26.3	179.0	33.6	50.0	17.3	50.0	19.8	14.0	0.5	0.30	< 0.1
1.0	1.0	55.0	26.9	75.0	34.4	55.0	27.4	55.0	23.3	15.0	0.8	0.30	< 0.1
1.0	1.0	65.0	29.2	85.0	38.8	70.0	32.3	60.0	27.2	15.0	1.0	0.30	< 0.1
1.0	1.0	70.0	31.5	90.0	42.2	75.0	33.0	75.0	30.8	15.0	1.4	0.30	< 0.1
1.0	1.0	75.0	33.6	100.0	45.5	70.0	33.6	85.0	33.9	15.0	1.8	0.30	< 0.1
2.0	1.0	85.0	35.9	105.0	48.6	250.0	34.8	90.0	36.4	15.0	2.2	0.30	< 0.1
—	—	90.0	38.2	115.0	52.4	—	—	430.0	39.7	15.0	2.7	0.20	< 0.1
—	—	100.0	40.9	125.0	56.1	—	—	—	—	15.0	3.3	0.40	< 0.1
—	—	110.0	43.6	130.0	59.4	—	—	—	—	15.0	3.9	—	—
—	—	120.0	46.4	145.0	63.9	—	—	—	—	15.0	4.6	—	—
—	—	125.0	49.0	155.0	68.8	—	—	—	—	15.0	5.3	—	—
—	—	135.0	52.5	175.0	72.7	—	—	—	—	15.0	5.9	—	—
—	—	675.0	58.8	875.0	82.0	—	< 0.1	—	—	15.0	6.0	—	—
—	—	—	—	—	—	—	—	—	—	15.0	6.5	—	—
—	—	—	—	—	—	—	—	—	—	20.0	7.5	—	—
—	—	—	—	—	—	—	—	—	—	30.0	8.6	—	—
—	—	—	—	—	—	—	—	—	—	150.0	11.3	—	< 0.1

7.3 Unfolding

There are a few different ways to perform the unfolding. Two examples are the iterative D'Agostini method [169] and the TUnfold algorithm [170]. The D'Agostini method is motivated by Bayesian statistics, requiring knowledge of a prior, in this case the particle-level distributions. Iteration of the algorithm reduces bias from the prior, however it also makes error propagation much more complicated. Alternatively, the TUnfold algorithm uses a more intuitive, least squares method of estimating the true distribution. However, it can suffer from an amplification of statistical fluctuations of the data, introduced by the inversion of the, often ill-conditioned, migration matrix.

The unfolding performed in the measurement presented in this thesis uses the TUnfold algorithm, because of its intuitiveness. A schematic of the unfolding process is shown in Fig. 63. It shows the relation between the true distribution, \vec{x} , and the measured distribution, \vec{y} . The migration matrix, \mathbf{A} , describes the migrations from a bin of the true distribution into any of the reconstructed bins. The average expected count of events, \tilde{y} , differs from the observed event counts due to statistical fluctuations.

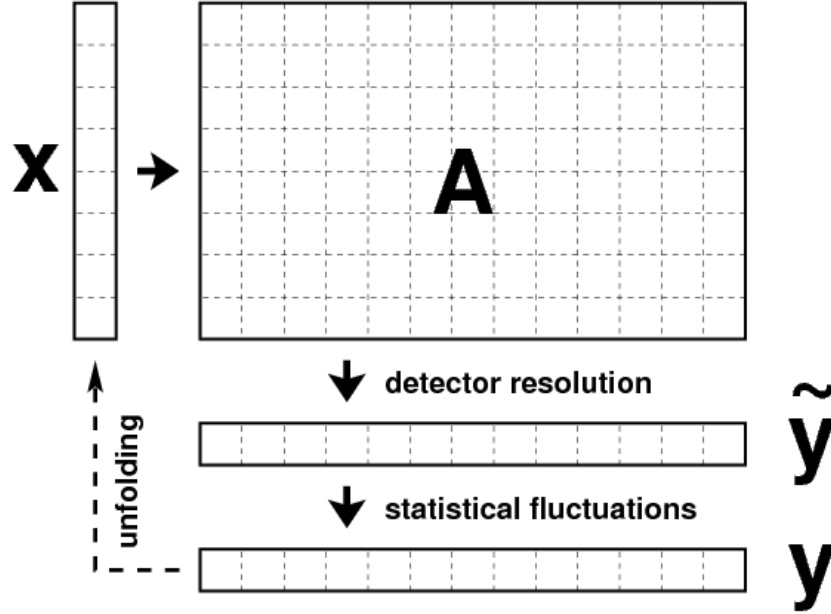


Figure 63: A schematic view of the migration effects and statistical fluctuations. Figure taken from [170].

The TUnfold algorithm uses a least squares method of estimating the true distribution, \vec{x} , from the reconstructed distribution, \vec{y} . Statistical fluctuations are smoothed using Tikhonov regularisation [171, 172]. The algorithm works to minimise

$$\mathcal{L}(x) = \mathcal{L}_1 + \mathcal{L}_2 \quad (7.3)$$

by finding the stationary point, where

$$\mathcal{L}_1 = (\vec{y} - \mathbf{A}\vec{x})^T \mathbf{V}_{yy}^{-1} (\vec{y} - \mathbf{A}\vec{x}) \quad (7.4)$$

and

$$\mathcal{L}_2 = \tau^2 (\vec{x} - \vec{x}_0)^T (\mathbf{L}^T \mathbf{L}) (\vec{x} - \vec{x}_0). \quad (7.5)$$

\mathcal{L}_1 represents the least squares minimisation and \mathcal{L}_2 describes the regularisation. \mathcal{L}_1 involves the measured observable distribution \vec{y} composed of m bins and the inverse of its covariance

matrix, \mathbf{V}_{yy}^{-1} . The covariance matrix is diagonal and holds the square of the statistical uncertainties in \vec{y} . The migration matrix \mathbf{A} is derived from the simulated POWHEG+PYTHIA sample.

\mathcal{L}_2 damps fluctuations in \vec{x} stemming from statistical fluctuations in \vec{y} . The parameter τ defines the strength of the regularisation. If it is too small then the unfolding is essentially unregularised, with the large negative correlations between neighbouring bins, introduced by the inversion of the migration matrix, enhancing the effect of statistical fluctuations in \vec{x} . If it is too large then the result will be biased towards the POWHEG+PYTHIA model, given by \vec{x}_0 . The regularisation can be applied by suppressing the differences in the size, derivative or curvature (second derivative) of $\vec{x} - \vec{x}_0$. When regularising by the derivative or curvature there are generally some positive correlations introduced between measurement bins. This is useful as it will cancel some of the negative correlations introduced by the unregularised unfolding term \mathcal{L}_1 , and for some value of τ the total correlation between bins of the measurement will be minimal. The choice of regularisation is given by the \mathbf{L} matrix. This thesis regularises by curvature, approximated by $(x_{i+1} - x_i) - (x_i - x_{i-1})$ leading to an \mathbf{L} matrix of order $(m-2) \times m$, where non-zero elements are present in elements $L_{i,i} = 1$, $L_{i,i+1} = -2$, $L_{i,i+2} = 1$.

The regularisation strength parameter τ is calculated by minimising the average global correlation coefficient. The components of the global correlation coefficient, ρ_i , are taken from the covariance matrix \mathbf{V}_{xx}

$$\rho_i = \sqrt{1 - \frac{1}{(\mathbf{V}_{xx}^{-1})_{ii}(\mathbf{V}_{xx})_{ii}}} \quad (7.6)$$

and the average global correlation is defined by

$$\sum_i \frac{\rho_i}{n}, \quad (7.7)$$

where n is the number of bins at particle level. Scans of the average global correlation coefficient in the e+jets channel over a range of regularisation strengths are shown in Figs. 64 and 64, with the best τ highlighted at the minimum. Similarly, the best regularisation strength is shown in Figs. 64 and 64 of App. D for the μ +jets channel. These are the regularisation strengths applied in the unfolding of the reconstructed $t\bar{t}$ yields of the measurements presented in this thesis.

7.4 Cross checking the unfolding

The unfolding is checked to ensure that negligible bias is introduced or mistreatment of the uncertainties on the reconstructed data occurs. The checks are performed using the particle-level truth information.

7.4.1 Checking the uncertainties

The effect of unfolding on the transformation of statistical uncertainties from the reconstructed data to the unfolded data is checked by the distribution of pulls. A pull is defined as the ratio of the difference between the number of unfolded events and the true number of events in a bin to the uncertainty on the number of unfolded events

$$\text{Pull}^i = \frac{x_{\text{truth}}^i - x_{\text{unf}}^i}{\sigma_{\text{unf}}^i}. \quad (7.8)$$

A set of pseudo experiments can be used to generate a pull distribution and if the mean and width of the pull distribution are close to zero and one respectively, then the normalisation and

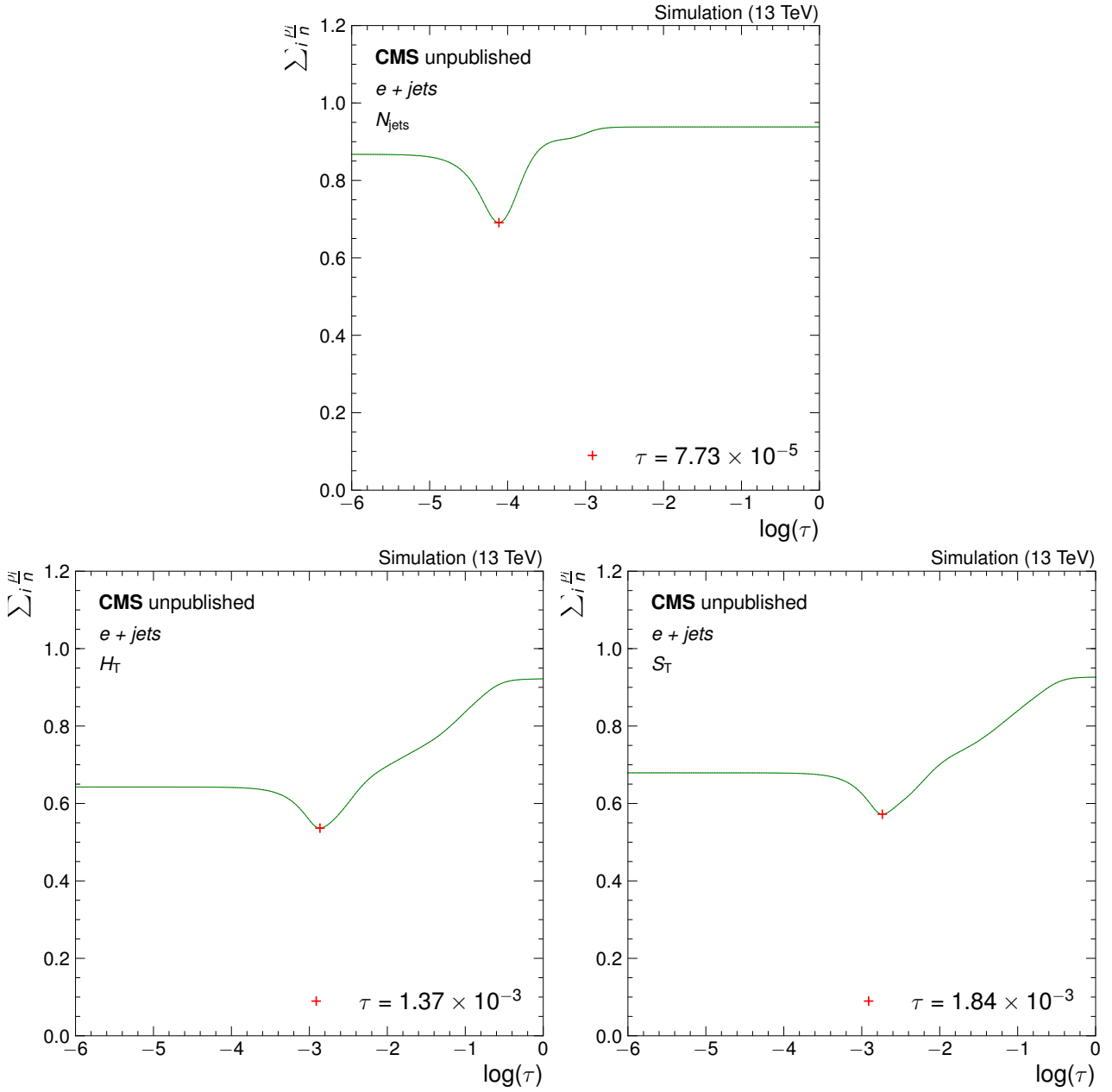


Figure 64: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the $e+jets$ channel, for the N_{jets} , H_T and S_T event variables. The optimal τ is shown as a red point at the minimum of the scan.

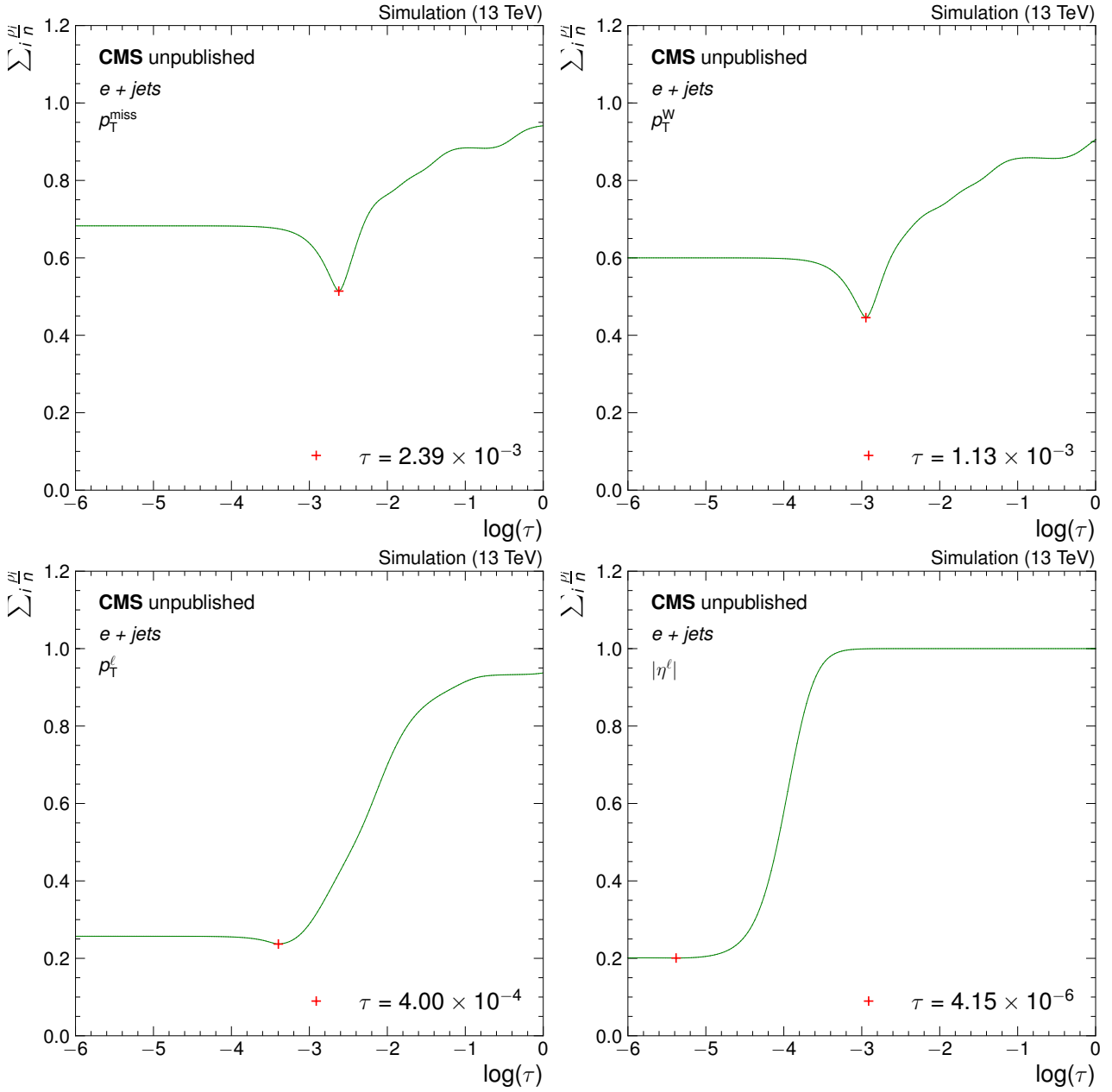


Figure 65: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the $e + jets$ channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables. The optimal τ is shown as a red point at the minimum of the scan.

statistical uncertainties of the unfolded distributions are treated correctly by the unfolding. Each pseudo experiment unfolds a new reconstructed $t\bar{t}$ yield with a new migration matrix and calculates the pull. The new $t\bar{t}$ yields and migration matrices are generated by throwing, per bin, a Poisson distributed random number with an average equal to those found in the corresponding bin of the POWHEG+PYTHIA simulation. A set of 5000 pseudo experiments are generated in total to produce the pull distributions. An example is shown in Fig. 66 for the H_T event variable in the e +jets channel. It can clearly be seen that the mean and width of the pull distribution are close to zero and one, so the unfolding is treating the uncertainty and normalisation correctly. The same conclusion is drawn for the other kinematic event variables, as shown in Figs. 67 and 68, where the means and widths are shown per bin and similarly in the μ +jets channel, shown in Figs. 105 and 106, of App. E.

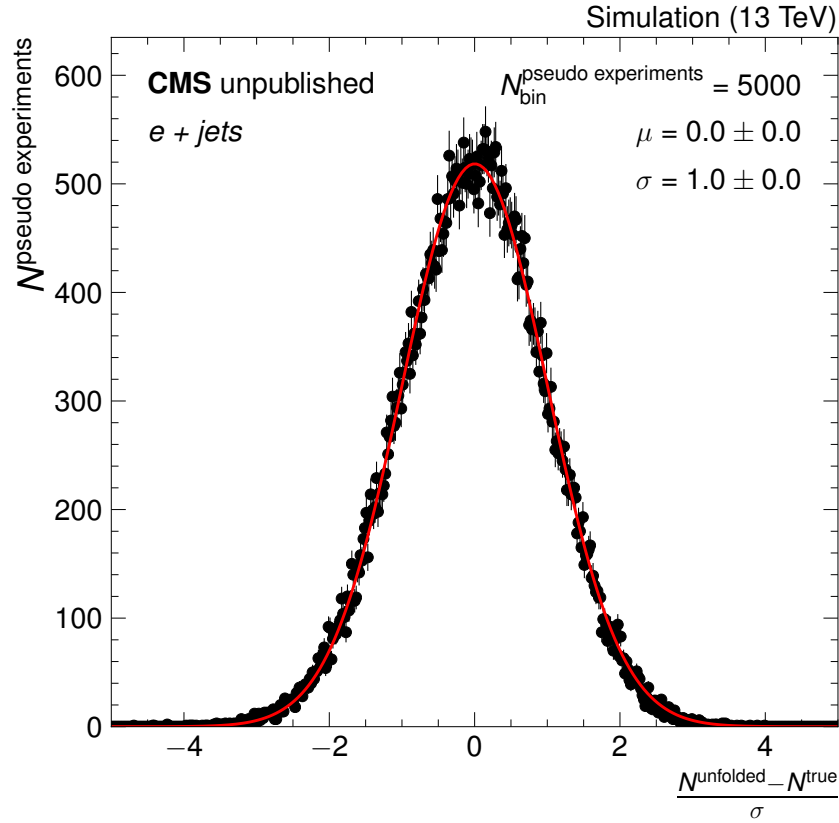


Figure 66: The pull distribution combining the pulls from all bins for 5000 pseudo experiments for the H_T variable in the e +jets channel.

7.4.2 Checking for bias

The migration matrices are model dependent because they are constructed directly from a simulated $t\bar{t}$ model. A model which poorly describes the data could introduce a bias in the unfolded distributions. In addition, if a measurement is regularised too much, then a bias towards the simulated $t\bar{t}$ model can be introduced. To test the size of any bias that could be introduced, each distribution is reweighted up and down such that they form an envelope covering any differences seen between data and the simulated POWHEG+PYTHIA sample. By unfolding the reweighted distributions, using a migration matrix derived from an identical, statistically independent POWHEG+PYTHIA sample, and then comparing to the true particle-level reweighted distributions, any bias that has been introduced will be revealed.

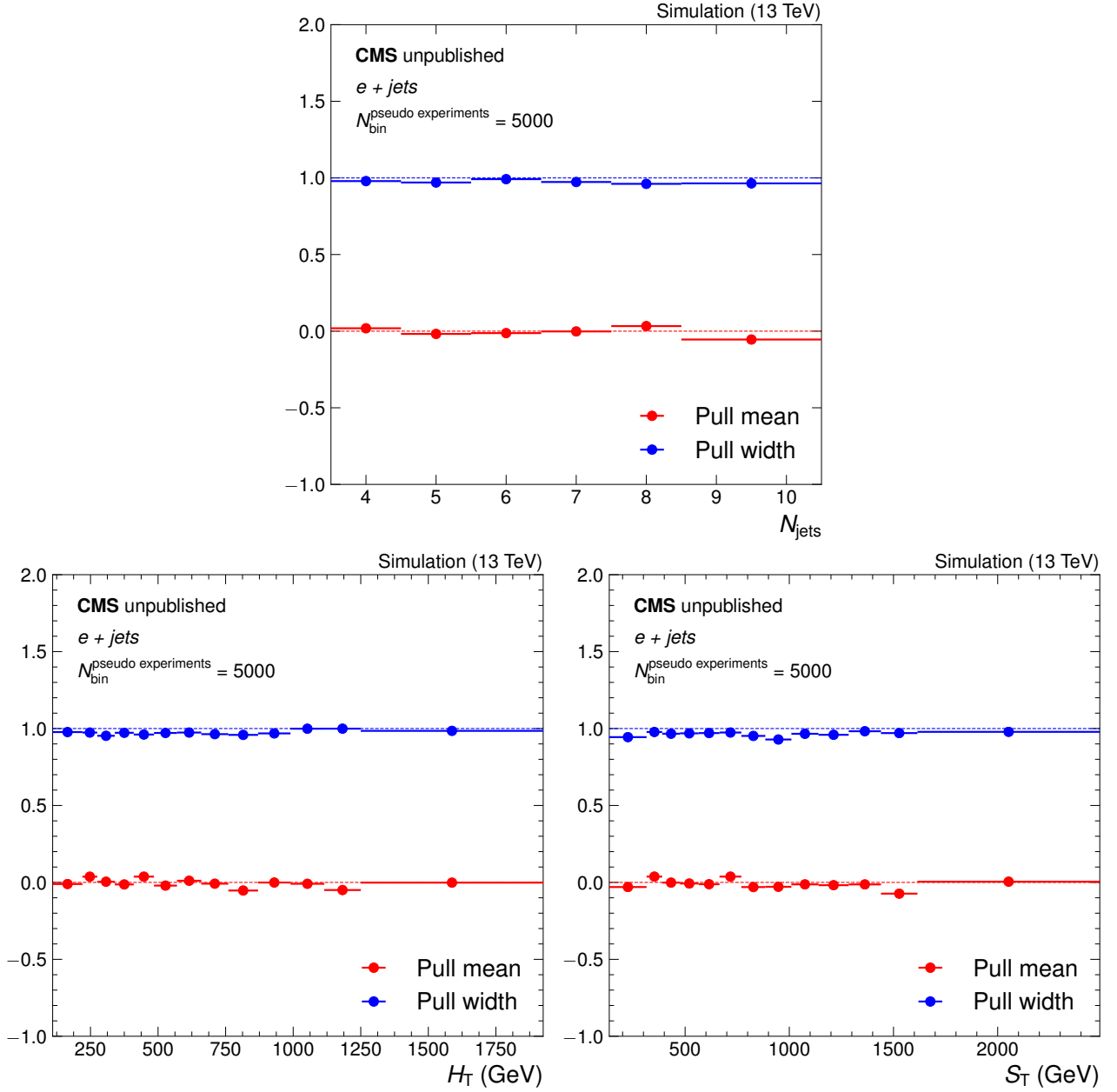


Figure 67: The pull mean and widths in relation to the bin numbers of the N_{jets} , H_T and S_T event variables in the $e+jets$ channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA simulation.

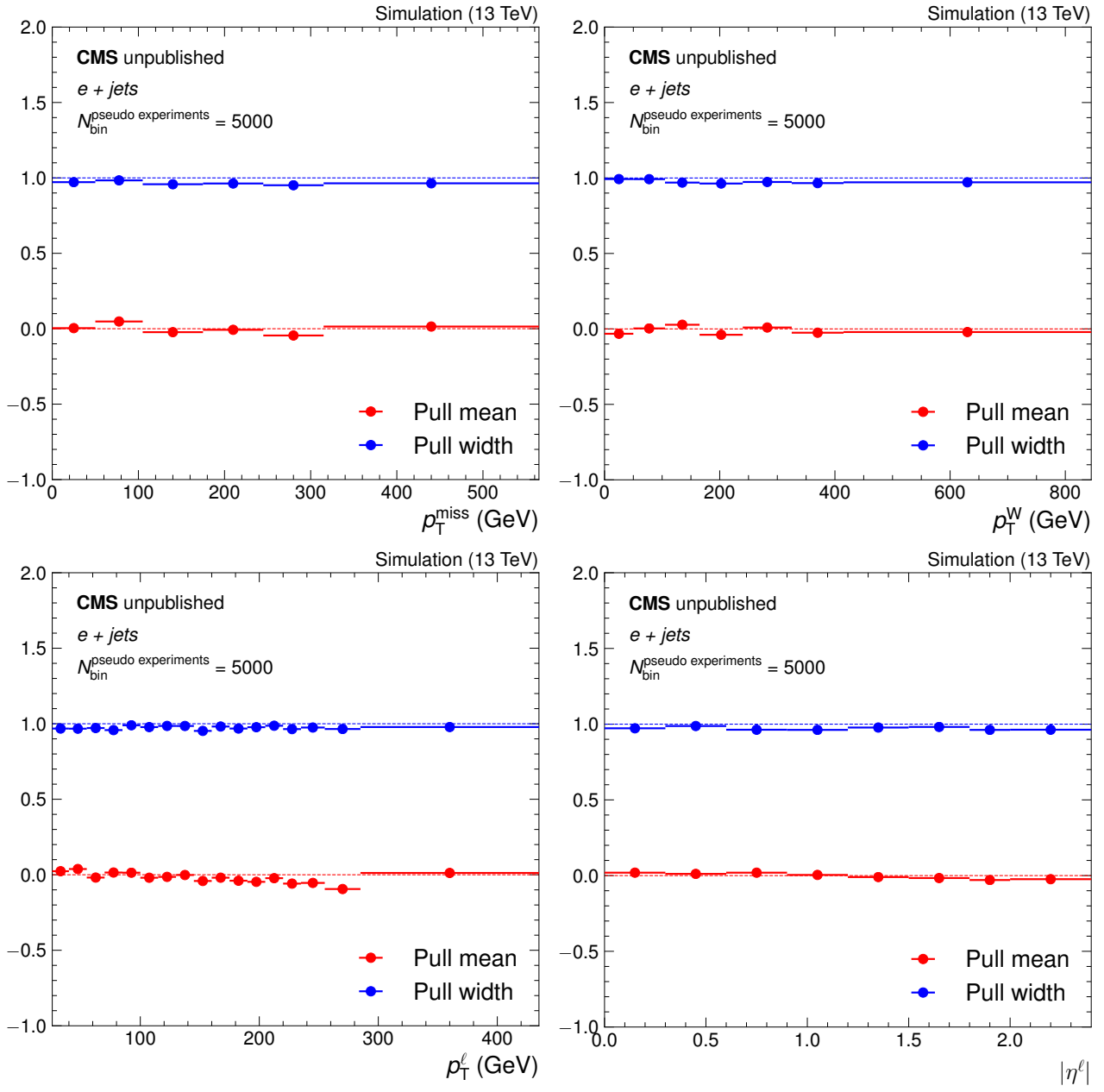


Figure 68: The pull mean and widths in relation to the bin numbers of the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the $e+jets$ channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA simulation.

The envelope covering the data-simulation differences is achieved by reweighting the particle-level p_T^{top} spectrum according to

$$w = 1 + (p_T^{\text{top}} \pm 100) \times 0.001. \quad (7.9)$$

Figure 69 shows an example of the envelope in the e+jets channel for the H_T event variable, which does indeed cover the data-simulation differences. The full sets of the reweighted distributions are given in App. F.

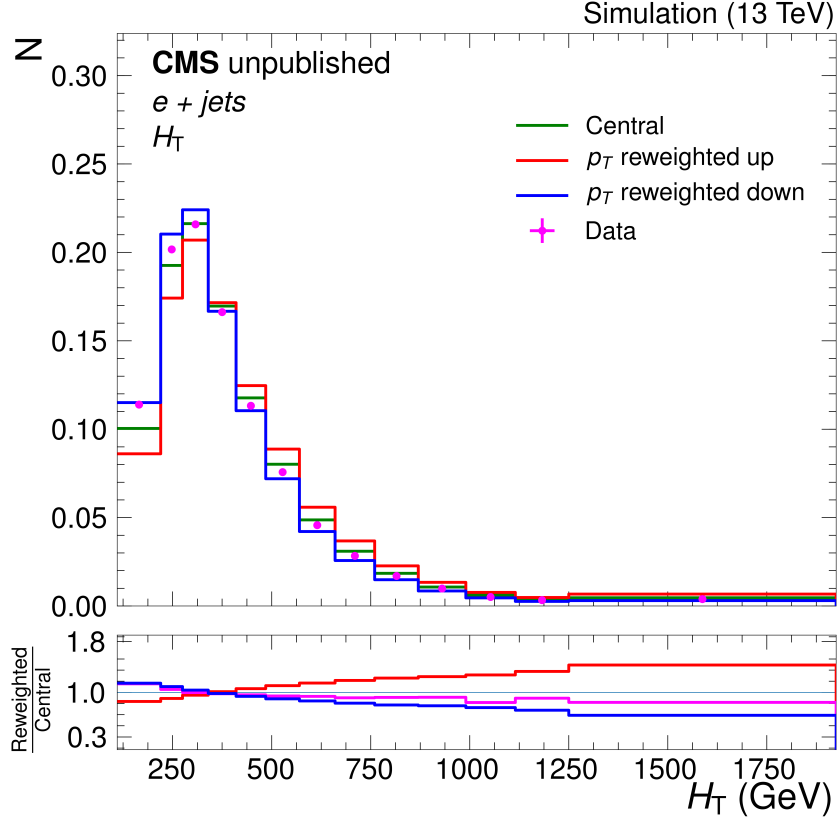


Figure 69: The H_T event distribution of the POWHEG+PYTHIA sample with the top quark p_T reweighted up and down to cover differences to data in the e+jets channel. The distributions are normalised to one.

The bias is defined as the ratio of unfolded differential cross section, calculated using the reweighted model $t\bar{t}$ yields, to the true model differential cross section. The calculation of the differential cross sections are defined in Sec. 7.4.3. Figures 70 and 71 show the unfolded model cross sections for the reweighted distributions compared against the true reweighted model cross sections in the e+jets channel. They are similarly shown in the μ +jets channel, in Figs. 111 and 112, of App. G. Bias will be shown by pulling the unfolded cross sections away from the true reweighted model cross sections. The ratios show the bias introduced by using the POWHEG+PYTHIA model in the migration matrix, compared to the total systematic uncertainty shown as the grey band. The systematic uncertainties are described in detail in Ch. 8. Any bias seen is small compared to the total systematic uncertainty for the reweighted samples. Bias measurements which have been calculated by unfolding alternative $t\bar{t}$ production models with the POWHEG+PYTHIA model are shown in Figs. 113, 114, 115 and 116 of App. G.

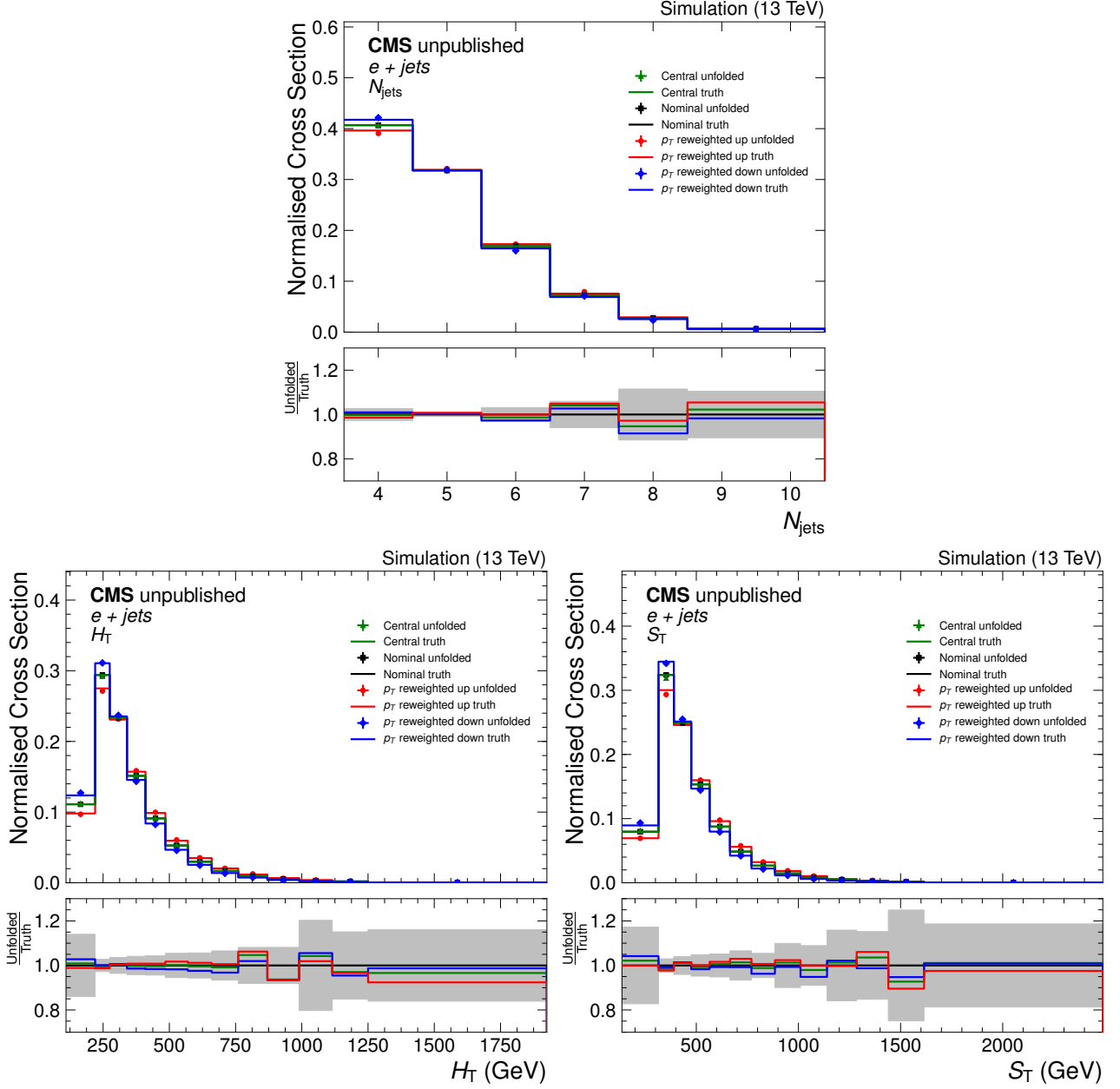
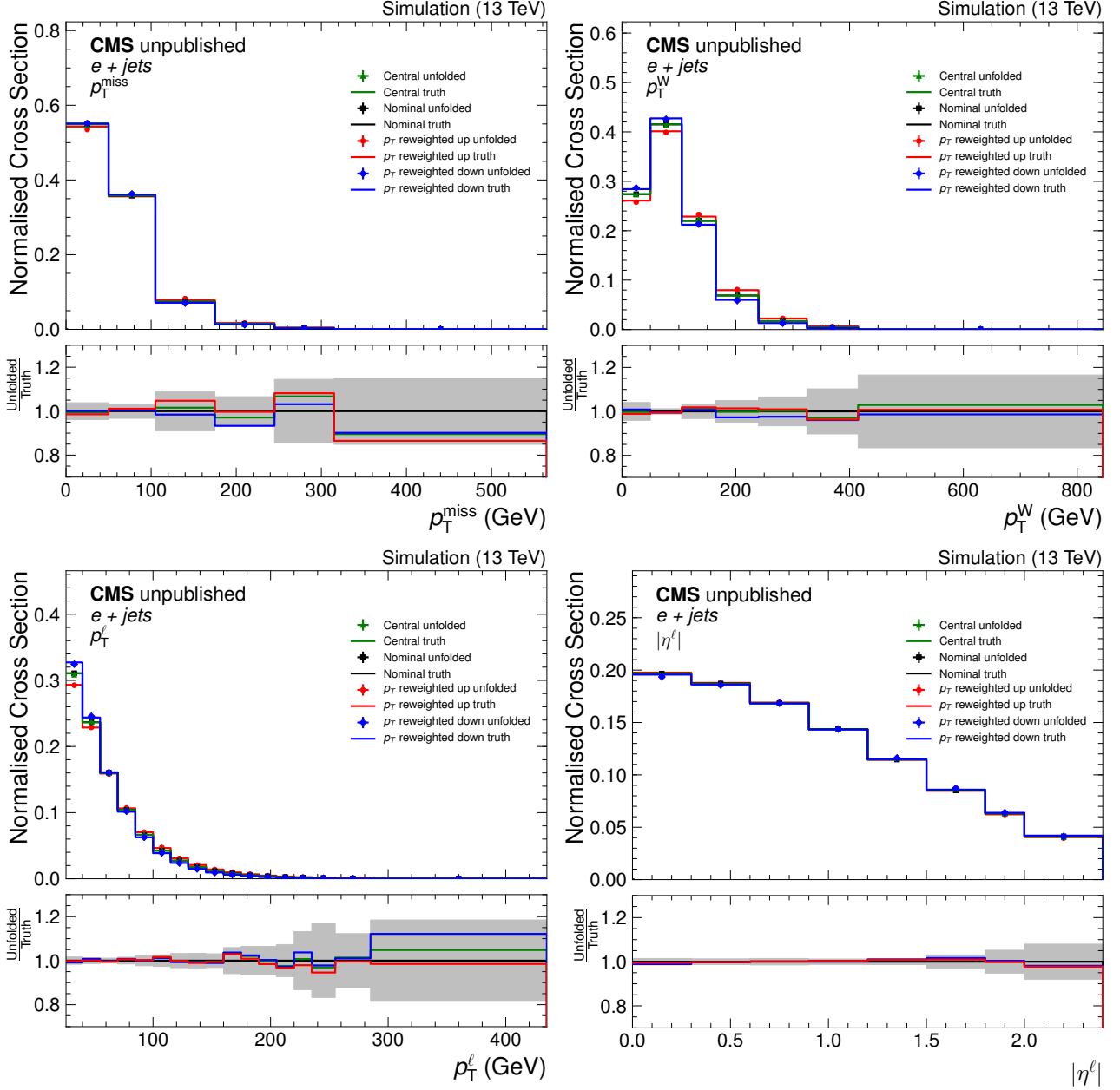


Figure 70: The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the $e + \text{jets}$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.



7.4.3 Calculating the cross sections

The yield of $t\bar{t}$ events are unfolded separately for the electron and muon channels to give the total number of events at particle level in the common visible phase space. The unfolded $t\bar{t}$ yields and their associated statistical uncertainties are added together in quadrature to form the combined channel, $e, \mu + \text{jets}$.

The absolute differential $t\bar{t}$ production cross section with respect to kinematic event variable, X , can be calculated from the number of unfolded $t\bar{t}$ events by

$$\frac{d\sigma_{t\bar{t}}^i}{dX} = \frac{N_{t\bar{t}}^i}{L\Delta X^i}. \quad (7.10)$$

It depends on the number of unfolded $t\bar{t}$ events in bin i , $N_{t\bar{t}}^i$, the width of bin i , ΔX^i , and the integrated luminosity of the data, L . Measuring the absolute differential cross sections provides maximal information about the cross sections and their associated uncertainties, however some uncertainties are highly correlated between bins of the measurement and can therefore be reduced significantly by normalising the differential cross sections. This allows for the shape of the differential cross sections to be investigated more closely. The normalised differential cross section is defined as

$$\frac{1}{\sigma_{t\bar{t}}^{\text{vis}}} \frac{d\sigma_{t\bar{t}}^i}{dX} = \frac{L}{\sum_j N_{t\bar{t}}^j} \frac{N_{t\bar{t}}^i}{L\Delta X^i} = \frac{1}{\sum_j N_{t\bar{t}}^j} \frac{N_{t\bar{t}}^i}{\Delta X^i}, \quad (7.11)$$

where $d\sigma_{t\bar{t}}^i/dX$ has been normalised by the total $t\bar{t}$ production cross section in the visible phase space at particle level $\sigma_{t\bar{t}}^{\text{vis}}$.

Before the measurements can be presented, a good understanding of the uncertainties on the differential cross section measurements is needed. The treatment of the uncertainties is explained in Ch. 8.

CHAPTER 8

Cross section uncertainties

When performing the cross section measurements, there are many factors which can affect the precision of the measurement. These effects are collectively known as uncertainties. They are subdivided into statistical and systematic uncertainties, where the statistical uncertainty is reducible by collecting more data and the systematic uncertainty by improvements in theoretical or experimental knowledge. Many searches for rare SM processes are dominated by the statistical uncertainty, however due to the multitude of $t\bar{t}$ pairs produced, the differential cross sections measured in this thesis are dominated by the systematic uncertainties.

8.1 Statistical uncertainties

The statistical uncertainties measured in the bins of each event variable in each channel are typically very small, $< 0.5\%$. Larger statistical uncertainties, $\approx 5\%$, can be found in the final bin of the differential cross section measurements where the largest is with respect to the p_T^{miss} event variable at 5.8% .

The statistical uncertainty from the finite size of the simulated POWHEG+PYTHIA sample used in the construction of the migration matrix for unfolding is also found to be very small, with no contribution larger than 2.2% .

8.2 Systematic uncertainties

The systematic uncertainties are evaluated and propagated to the final result by recalculating the migration matrix using a modified POWHEG+PYTHIA simulation and/or by modifying the background predictions. The unfolded yields in the $e+\text{jets}$ and $\mu+\text{jets}$ channels are then combined into the $e, \mu+\text{jets}$ channel for each systematic variation, before the differential cross sections are calculated. The uncertainty due to a particular systematic is taken as the difference between the cross sections estimated from the systematic variation and the nominal model. Specifically, it is calculated from the variance of the systematic uncertainty, implying an inherent symmetrisation of each uncertainty. If the systematic uncertainty is two sided, i.e. contains upward and downward variations, then it is symmetrised according to the average of

the variations. This symmetrisation of the uncertainties substantially simplifies the process of error propagation. The inherent inaccuracies in this process are negligible.

The systematic uncertainties can be split into two discrete categories. The first category contains the experimental uncertainties, which represent our limited understanding of the detector performance. The second is the set of theoretical uncertainties related to the modelling of the $t\bar{t}$ event. They are discussed in more depth in the following sections.

8.2.1 Experimental uncertainties

The systematic uncertainties introduced by other experimental measurements, for example those used in the calculation of scale factors, cross sections or other experimental quantities, stem from the limited precision of those measurements.

Uncertainty in the luminosity

When the integrated luminosity is changed, the number of events in each bin and hence the $t\bar{t}$ yield will change. The event yields are reweighted up and down according to the uncertainty of the integrated luminosity of the data, which was measured to be $\pm 2.5\%$ [173]. This uncertainty is seen as 2.5% in the case of the absolute differential cross sections and is completely normalised away in the normalised differential cross section measurements.

Uncertainty in the number of inelastic interactions

The uncertainty in the number of additional inelastic interactions in the same or nearby bunch crossings is measured by varying the total inelastic cross section which is used in the calculation of the pileup distribution as seen in Sec. 6.2, by its uncertainty of $\pm 4.6\%$ [109]. The uncertainty is typically very small at $< 0.5\%$.

Uncertainty in b tagging efficiency

The uncertainty of the b jet identification and tagging efficiency in the simulation is taken from the uncertainty in the b tag scale factors, dependent on p_T , η and quark flavour [162]. The contribution from light jets is calculated independently to that from the b/c jets. The pairs of up and down variations are added in quadrature to give a total uncertainty. The uncertainty in the b tagging efficiency is typically negligible for normalised differential cross sections, however can be much bigger, $\approx 3 - 5\%$, in the absolute differential cross section measurements.

Uncertainty in the lepton scale factors

There is an uncertainty present in the lepton correction factors used in the measurements presented in this thesis. The uncertainty is estimated by simultaneously varying the scale factors for the lepton identification, isolation, reconstruction and trigger efficiencies up and down by their uncertainties. The uncertainties are small, typically $< 3\%$.

Uncertainty in the jet energy corrections

The uncertainties in the JES and JER corrections are estimated as functions of jet p_T and η [161], by varying the scale factors by their uncertainties. As the corrections directly effect the event selection, the variations must be applied before the event selection and propagated through the measurements, as opposed to changing an event weight. It is, perhaps, unsurprising that the uncertainties in the JEC, particularly from the JES, are the largest contributions to the

total uncertainty for the variables relating to the hadronic activity of the event. The uncertainty from the JES is typically $\approx 5\%$, but can exceed 15% in the first bin of the S_T event variable.

In addition to the uncertainties of the jets in the event, the p_T^{miss} variable also depends on the uncertainties of the p_T of the reconstructed electron, muon, tau lepton and unclustered PF candidates. The uncertainties on these PF candidates are propagated to the event variables dependent on p_T^{miss} and are found to have a very small effect only. The uncertainties are typically 1-2% with the largest coming from the misreconstruction of the muon p_T on the p_T^{miss} variable at 3.5% in one bin.

Uncertainty in the background cross sections

Variations in the background cross sections have a direct effect on the yield of $t\bar{t}$ events. The uncertainties in the cross sections for the single top quark and vector boson backgrounds are based on measurements performed in [174–176]. These measurements are performed in phase spaces very different from that of the measurements presented in this thesis and so the uncertainties have been inflated from $\pm 13\%$ and $\pm 20\%$ to $\pm 30\%$ and $\pm 50\%$ respectively. The uncertainties from these backgrounds are typically small, $< 5\%$, however in the final bins of the p_T^ℓ and p_T^W event variables the uncertainty in the single top cross section increases to around 7 – 8%.

The uncertainty in the shape and normalisation of the multijet QCD is taken by using the shape and normalisation prediction from the alternative QCD control region. This can lead to uncertainties of up to $\pm 30\%$ and $\pm 60\%$ respectively in any one bin. The uncertainty from the prediction of the muon multijet QCD is negligible across all bins of all event variables, $< 1\%$, however the uncertainty from the prediction of the electron multijet QCD is large at high $|\eta^\ell|$.

8.2.2 Modelling uncertainties

The systematic uncertainties in the modelling of the $t\bar{t}$ events are given by variations in a complete set of associated parameters incorporated in the POWHEG+PYTHIA simulation. These can either take the form of additional POWHEG+PYTHIA simulations, stated in Tab. 5 or from additional sets of weights stored in the main POWHEG+PYTHIA simulation.

Uncertainty in the matrix-element and parton shower scales

The uncertainties in the modelling of the hard process are quantified through the matrix-element energy scales: the factorisation and renormalisation scales. It is estimated by varying the scales independently by a factor of two, up and down, and, additionally, varying the scales simultaneously by the same factors. The choice of the factor of two is arbitrary, however, it is seen as a sensible choice, because a variation which is too large can mean that the factorisation and renormalisation scales become unrealistic.

As well as the uncertainties due from the energy scales in the matrix-element, the modelling of radiation in the parton shower is also sensitive to the energy scale at which it is emitted, μ . The uncertainty from μ in the ISR modelling is estimated by varying it up and down by a factor of two. Similarly, the μ in for the FSR modelling is varied up and down by a factor of $\sqrt{2}$. The variation is reduced to $\sqrt{2}$ by constraints measured from the LEP collider [177].

The energy scales uncertainty is defined as the envelope of matrix-element and parton shower scale uncertainties. The dominant component in the energy scales uncertainty originates from the final-state radiation modelling. The uncertainty is large in the absolute differential measurements, $\approx 4 - 6\%$, but much reduced by the normalisation, in the normalised differential measurements.

Uncertainty in the matching

The matching of jets in the matrix-element to the parton shower, for the POWHEG+PYTHIA model, was introduced in Sec 4.3. The h_{damp} parameter is used to regulate the number of high- p_T soft emissions. The uncertainty in the matching between the matrix-element and parton shower is estimated by varying this h_{damp} parameter within its uncertainty, $1.58^{+0.66}_{-0.59} \times m_t$ [82]. The uncertainty is small in comparison to other modelling uncertainties at 1 – 2%.

Uncertainty in the underlying event

The choice of the underlying event tune can have an impact on the measurements presented in this thesis, because it is responsible for the number of MPIs and CRs that can happen in an event. The uncertainty is estimated by varying up and down a number of tuned parameters of the underlying event by their uncertainties. Details on the underlying event parameters can be found in [82]. The uncertainty is of a comparable size to the matching uncertainty.

Uncertainty in the b quark fragmentation

The uncertainty in the transfer of momentum from b quarks to B hadrons is estimated by varying the tuned parameter $x_b = p_T^B/p_T^{\text{bjet}}$ for each tagged particle level b jet up and down by its uncertainty. The variables p_T^B and p_T^{bjet} represent the transverse momenta of the B hadron and the particle-level b jet respectively. The uncertainty is denoted as the fragmentation uncertainty and is typically < 1%. An additional fragmentation uncertainty is also included by using the difference to an alternative fragmentation model (Peterson model [178]), which is slightly larger at 1 – 3% for the absolute cross sections.

Uncertainty in the single-lepton branching fractions of B hadrons

As well as the b quark fragmentation, the b jet energy response is sensitive to the single-lepton branching fractions of B hadrons. The systematic uncertainty introduced by the choice of branching fractions used in the POWHEG+PYTHIA model is estimated by reweighting the fractions to those reported in the 2016 [12]. These uncertainties are negligible at < 0.3%.

Uncertainty in the CR

The uncertainty in the modelling of the CR of the event is estimated by comparing the cross sections produced when including and excluding the colour reconnection on the top quark decay products (Early resonance decays). Two additional uncertainties are included by using the two different colour reconnection models discussed in Sec. 4.4.1, which are the QCD-based and Gluon move models. Each model is currently treated as a separate systematic uncertainty. Perhaps a better way to estimate the uncertainty would be to take the envelope of all the models. Figure 72 shows, for the absolute differential cross section measurement with respect to the H_T event variable, an example comparison of the total uncertainty from the modelling of CR that has been calculated both by adding the uncertainty from each model in quadrature and also by taking the envelope of each uncertainty. The difference in the relative uncertainties between the two predictions is of the order 2%, which is acceptable. Similar differences are also seen in relation to the other event variables. Overall, the CR can introduce a large systematic uncertainty in the differential cross section measurements, typically between 2 – 5%, but exceeding 10% in the case of one bin of the measurement with respect to the S_T event variable.

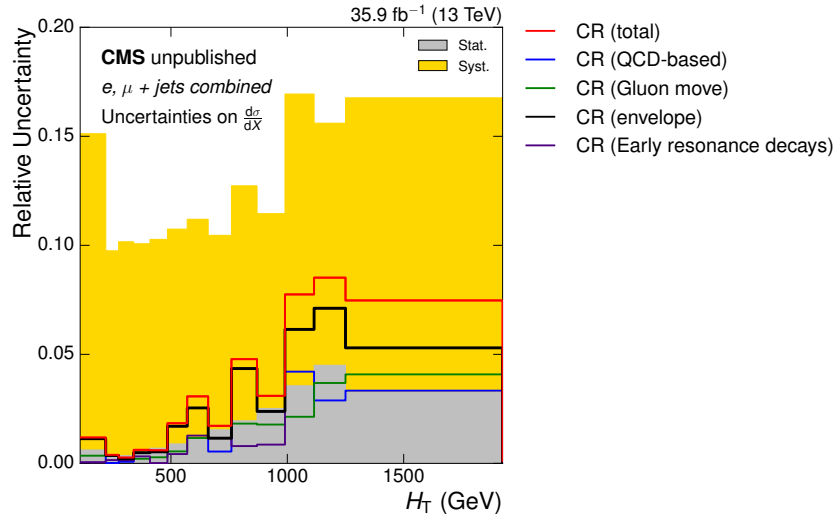


Figure 72: The colour reconnection uncertainties for the absolute differential cross section measurement as a function of the H_T event variable. Overlaid is the total colour reconnection uncertainty calculated both by combining in quadrature and taking the envelope of the three colour reconnection models. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

Uncertainty in the PDF

The NNPDF30_nlo_as_0118 PDF set used in this analysis defines the initial kinematic properties of the colliding partons used in the POWHEG+PYTHIA simulation of the $t\bar{t}$ events and therefore can have an impact on the yields of the $t\bar{t}$ measured and the shape of the differential cross sections. The uncertainty is estimated by considering 100 independent replicas. The RMS of the uncertainties, taken as the variations produced from each replica, is defined as the PDF uncertainty. The uncertainty from the choice of $\alpha_S = 0.118$ used in the NNPDF30_nlo_as_0118 set is estimated by varying α_S by ± 0.001 which is then combined in quadrature with the PDF uncertainty. The PDF uncertainty is negligible with no uncertainty larger than 0.6% in any bin.

Uncertainty in the top mass

The impact of the top quark mass on the $t\bar{t}$ signal selection efficiency is estimated by variations from two independent samples simulated with $m_t = 169.5$ GeV and 175.5 GeV. The variations are then linearly scaled down to be comparable with the world top quark mass uncertainty of ± 1 GeV given in the 2016 PDG [12]. This results in a systematic uncertainty of $\approx 1 - 2\%$.

Uncertainty in the top p_T

The effect of the top quark p_T has been seen and discussed in Sec 6.6.1. The uncertainty originating from the mismodelling of the top quark p_T spectrum is estimated by reweighting the p_T distribution to that measured by [87, 88], using Eq. 6.7. The reweighting can cause the yield of $t\bar{t}$ events to be modified by up to 20% in any one bin, however the total number of $t\bar{t}$ events remains equivalent. The p_T^{top} reweighting results in a very small uncertainty at $< 1\%$.

8.3 Summarising the uncertainties

A complete breakdown detailing the smallest and largest contributions over all bins of each systematic uncertainty for normalised cross section measurements is shown in Tab. 13 and for absolute cross section measurements in Tab. 14. Similar values for the total statistical uncertainty, total systematic uncertainty and the total uncertainty are also shown. Sources of uncertainty that only affect $p_{\text{T}}^{\text{miss}}$ are indicated by — for other variables. The typical uncertainty for the normalised cross section measurements is below 5% but can be as large as 18%. For the measurements of the absolute cross section the uncertainties are typically 10% but can be as large as 21% in the tails of the distributions. Figures 117, 118, 119 and 120 of App. H show the largest relative systematic uncertainties in every bin for each event variable.

In the systematic uncertainties for the normalised cross section measurements, the dominant source of uncertainty for the hadronic based variables is related to the JES. For the measurement with respect to the $|\eta^\ell|$ event variable the normalisation of the background QCD originating from the e+jets channel dominates at high pseudorapidity, however the total uncertainty is small typically at 0.5%. There is no specific systematic uncertainty which dominates in the measurements with respect to p_{T}^ℓ .

For the absolute cross section measurements the JES is still among the dominant uncertainties present, however more significant contributions are present from the $t\bar{t}$ modelling uncertainties which are fundamentally reduced in the normalised cross section measurements. In particular, the uncertainty originating from the matrix-element and parton shower energy scales is large. The b quark tagging efficiency is also large at approximately 3% in all bins.

Table 13: The upper and lower bounds, in %, from each source of systematic uncertainty in the normalised differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown.

Relative uncertainty source (%)	N_{jets}	H_T	S_T	p_T^{miss}	p_T^W	p_T^ℓ	$ \eta^\ell $
b tagging efficiency	0.1 – 0.8	0.2 – 1.1	0.2 – 1.5	0.1 – 1.2	0.1 – 1.7	0.1 – 1.9	0.1 – 0.5
Electron efficiency	0.1 – 0.2	0.1 – 0.7	0.1 – 0.9	0.1 – 0.7	0.1 – 1.4	0.3 – 2.1	0.1 – 0.8
Muon efficiency	0.1 – 0.3	0.1 – 0.2	0.1 – 0.3	0.1 – 0.2	0.1 – 0.6	0.1 – 1.0	0.1 – 0.1
JER	0.1 – 0.6	0.1 – 0.7	0.2 – 1.8	0.6 – 5.8	0.2 – 2.1	0.1 – 0.2	<0.1
JES	0.1 – 5.5	2.1 – 13.6	2.1 – 15.9	2.1 – 7.1	0.5 – 4.9	0.1 – 2.0	0.1 – 0.2
Electron transverse momentum in p_T^{miss}	—	—	0.1 – 0.3	0.1 – 0.9	0.1 – 0.6	—	—
Muon transverse momentum in p_T^{miss}	—	—	0.1 – 0.9	0.1 – 3.5	0.1 – 0.9	—	—
Tau transverse momentum in p_T^{miss}	—	—	0.1 – 1.4	0.1 – 1.2	0.1 – 1.4	—	—
Unclustered transverse momentum in p_T^{miss}	—	—	0.1 – 1.7	0.2 – 1.9	0.1 – 1.0	—	—
Electron QCD bkg cross section	0.1 – 0.5	0.1 – 0.9	0.1 – 1.5	0.2 – 0.6	0.1 – 0.8	0.1 – 4.5	0.2 – 2.9
Muon QCD bkg cross section	0.1 – 0.2	0.1 – 0.5	0.1 – 0.8	0.1 – 0.2	0.1 – 0.2	0.1 – 0.2	0.1 – 0.1
Electron QCD bkg shape	<0.1	0.1 – 0.4	0.1 – 1.0	0.1 – 0.1	0.1 – 1.5	0.1 – 4.7	0.1 – 1.5
Muon QCD bkg shape	0.1 – 0.1	0.1 – 0.7	0.1 – 0.7	<0.1	<0.1	0.1 – 0.1	0.1 – 0.1
Single top quark cross section	0.1 – 0.4	0.1 – 2.1	0.1 – 4.4	0.1 – 4.9	0.1 – 7.1	0.1 – 6.0	<0.1
V+jets cross section	0.1 – 0.3	0.1 – 2.5	0.1 – 3.7	0.1 – 2.0	0.1 – 3.6	0.1 – 5.4	0.1 – 1.5
PDF	0.1 – 0.3	0.1 – 0.3	0.1 – 0.6	0.1 – 0.3	0.1 – 0.4	0.1 – 0.4	<0.1
Color reconnection (Gluon move)	0.1 – 2.8	0.1 – 4.0	0.1 – 11.7	0.2 – 0.9	0.1 – 1.0	0.2 – 4.8	0.1 – 0.4
Color reconnection (QCD-based)	0.1 – 2.0	0.1 – 4.2	0.1 – 6.6	0.4 – 4.2	0.1 – 3.5	0.1 – 7.6	0.1 – 1.2
Color reconnection (Early resonance decays)	0.2 – 3.9	0.1 – 7.1	0.1 – 4.1	0.1 – 1.6	0.1 – 3.8	0.1 – 5.0	0.1 – 1.0
Fragmentation	0.1 – 0.4	0.1 – 0.5	0.1 – 0.5	0.1 – 0.6	0.1 – 0.6	0.1 – 0.4	<0.1
h_{damp}	0.3 – 3.8	0.1 – 3.1	0.2 – 2.9	0.1 – 2.3	0.1 – 2.7	0.1 – 2.8	0.2 – 1.2
Top quark mass	0.2 – 1.0	0.1 – 3.1	0.2 – 3.5	0.1 – 4.0	0.2 – 1.1	0.2 – 4.5	0.1 – 0.6
Peterson fragmentation model	0.1 – 1.3	0.1 – 0.6	0.1 – 0.9	0.1 – 1.1	0.1 – 1.0	0.1 – 1.3	<0.1
Energy scales	0.4 – 4.3	0.5 – 4.5	0.5 – 4.9	0.2 – 2.4	0.3 – 3.5	0.1 – 4.5	0.1 – 0.7
hadron decay semileptonic branching fraction	0.1 – 0.1	0.1 – 0.1	0.1 – 0.1	<0.1	<0.1	<0.1	<0.1
Top quark p_T	0.1 – 0.7	0.1 – 0.9	0.1 – 1.0	0.1 – 0.8	0.1 – 0.9	0.1 – 1.3	<0.1
Underlying event tune	0.1 – 2.7	0.1 – 5.5	0.2 – 4.4	0.1 – 5.4	0.2 – 2.6	0.2 – 6.1	0.1 – 0.9
Simulated sample size	0.1 – 1.6	0.1 – 1.6	0.1 – 1.9	0.1 – 2.2	0.1 – 1.4	0.1 – 1.7	0.1 – 0.4
Additional interactions	0.1 – 0.4	0.1 – 1.0	0.1 – 1.7	0.1 – 1.5	0.1 – 0.9	0.1 – 1.0	<0.1
Integrated luminosity	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Total systematic uncertainty	0.6 – 9.6	2.7 – 14.1	2.8 – 17.4	2.9 – 11.7	0.8 – 12.6	0.7 – 13.4	0.7 – 4.4
Total statistical uncertainty	0.3 – 3.9	0.3 – 4.5	0.3 – 5.3	0.2 – 5.8	0.2 – 4.1	0.3 – 4.7	0.3 – 1.0
Total uncertainty	0.7 – 10.0	2.8 – 14.1	2.8 – 18.2	3.0 – 13.0	0.8 – 13.2	0.8 – 14.2	0.8 – 4.5

Table 14: The upper and lower bounds, in %, from each source of systematic uncertainty in the absolute differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown.

Relative uncertainty source (%)	N_{jets}	H_T	S_T	p_T^{miss}	p_T^W	p_T^ℓ	$ \eta^\ell $
b tagging efficiency	3.1 – 4.0	3.5 – 4.5	3.5 – 5.0	3.6 – 4.8	3.6 – 5.2	3.6 – 5.4	3.6 – 4.2
Electron efficiency	1.6 – 1.9	1.6 – 2.3	1.4 – 2.5	1.6 – 2.4	1.2 – 3.0	1.0 – 3.5	1.0 – 2.1
Muon efficiency	2.0 – 2.5	2.2 – 2.5	2.2 – 2.6	2.3 – 2.6	2.3 – 2.9	2.2 – 3.3	2.1 – 2.3
JER	0.1 – 0.8	0.1 – 1.1	0.2 – 2.3	0.4 – 5.6	0.4 – 1.6	0.1 – 0.3	0.3 – 0.4
JES	1.4 – 9.2	4.3 – 12.8	4.3 – 15.1	2.2 – 10.9	2.2 – 7.8	1.8 – 4.1	3.8 – 4.1
Electron transverse momentum in p_T^{miss}	—	—	0.1 – 0.3	0.1 – 0.9	0.1 – 0.6	—	—
Muon transverse momentum in p_T^{miss}	—	—	0.1 – 0.9	0.1 – 3.5	0.1 – 0.8	—	—
Tau transverse momentum in p_T^{miss}	—	—	0.1 – 1.4	0.1 – 1.2	0.1 – 1.4	—	—
Unclustered transverse momentum in p_T^{miss}	—	—	0.1 – 1.7	0.2 – 1.9	0.1 – 1.0	—	—
Electron QCD bkg cross section	0.1 – 0.8	0.2 – 1.4	0.2 – 2.1	0.1 – 0.8	0.1 – 1.3	0.3 – 5.0	0.1 – 3.6
Muon QCD bkg cross section	0.1 – 0.3	0.1 – 0.7	0.1 – 1.1	0.1 – 0.3	0.1 – 0.4	0.1 – 0.4	0.1 – 0.3
Electron QCD bkg shape	<0.1	0.1 – 0.3	0.1 – 1.0	0.1 – 0.1	0.1 – 1.6	0.1 – 4.8	0.1 – 1.5
Muon QCD bkg shape	0.1 – 0.1	0.1 – 0.7	0.1 – 0.8	<0.1	<0.1	0.1 – 0.1	0.1 – 0.1
Single top quark cross section	1.1 – 1.7	1.1 – 3.5	1.1 – 5.8	1.3 – 6.3	1.1 – 8.4	1.3 – 7.4	1.4 – 1.5
V+jets cross section	0.7 – 1.1	0.6 – 3.4	0.5 – 4.6	0.7 – 2.9	0.7 – 4.5	0.6 – 6.3	0.6 – 2.5
PDF	0.1 – 0.3	0.1 – 0.4	0.1 – 0.6	0.1 – 0.3	0.1 – 0.4	0.1 – 0.5	<0.1
Color reconnection (Gluon move)	0.2 – 2.8	0.2 – 4.1	0.1 – 11.8	0.2 – 1.0	0.1 – 1.1	0.2 – 4.7	0.1 – 0.5
Color reconnection (QCD-based)	0.2 – 2.1	0.1 – 4.3	0.1 – 6.7	0.3 – 4.6	0.3 – 3.8	0.1 – 7.9	0.1 – 1.7
Color reconnection (Early resonance decays)	0.1 – 3.9	0.1 – 7.1	0.1 – 4.1	0.1 – 1.4	0.1 – 3.8	0.1 – 5.0	0.1 – 1.1
Fragmentation	0.1 – 0.7	0.6 – 1.6	0.5 – 1.4	0.1 – 0.7	0.2 – 0.8	0.1 – 0.8	0.3 – 0.4
h_{damp}	0.1 – 3.5	0.2 – 3.1	0.1 – 3.0	0.6 – 2.1	0.2 – 3.2	0.4 – 3.4	0.5 – 2.0
Top quark mass	0.7 – 2.0	0.3 – 3.4	0.3 – 3.7	0.3 – 5.0	0.4 – 2.0	0.3 – 3.5	0.9 – 1.7
Peterson fragmentation model	0.3 – 2.0	1.6 – 2.7	1.9 – 3.2	1.1 – 2.6	1.3 – 2.6	1.2 – 2.9	1.5 – 1.5
Energy scales	2.7 – 6.5	2.6 – 6.4	3.0 – 7.0	4.0 – 6.6	4.6 – 6.2	3.9 – 6.2	4.9 – 5.6
hadron decay semileptonic branching fraction	0.2 – 0.3	0.1 – 0.3	0.1 – 0.3	0.2 – 0.3	0.2 – 0.3	0.2 – 0.3	0.2 – 0.3
Top quark p_T	0.4 – 1.2	0.1 – 0.8	0.1 – 0.9	0.1 – 1.5	0.1 – 1.0	0.1 – 1.0	0.6 – 0.7
Underlying event tune	0.1 – 2.9	0.2 – 5.3	0.2 – 4.4	0.2 – 5.4	0.1 – 2.6	0.1 – 6.0	0.2 – 0.9
Simulated sample size	0.1 – 1.6	0.1 – 1.6	0.1 – 1.9	0.1 – 2.2	0.1 – 1.4	0.1 – 1.7	0.1 – 0.4
Additional interactions	0.1 – 0.3	0.1 – 0.7	0.1 – 1.3	0.3 – 1.3	0.1 – 0.6	0.1 – 0.8	0.1 – 0.3
Integrated luminosity	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5
Total systematic uncertainty	8.7 – 13.4	9.7 – 15.9	9.5 – 20.0	8.8 – 17.1	8.3 – 17.5	8.6 – 16.1	8.8 – 10.6
Total statistical uncertainty	0.3 – 3.9	0.4 – 4.5	0.3 – 5.3	0.2 – 5.8	0.2 – 4.1	0.3 – 4.7	0.3 – 1.1
Total uncertainty	8.7 – 13.7	9.7 – 16.2	9.5 – 20.7	8.8 – 18.1	8.3 – 18.0	8.6 – 16.7	8.8 – 10.6

CHAPTER 9

Differential cross section results

Now that the differential cross section results are ready to be presented, it is worth stating a brief summary of the steps required to get to this stage.

Firstly, in order to produce $t\bar{t}$ events, a large collider is needed, in this case the LHC, and a detector capable of producing excellent data, the CMS experiment. The detector needs to be granular enough to reconstruct the energy and momentum of each particle in the event. The LHC and the CMS detector produce a large amount of good quality data.

In order to compare to the data, a prediction is needed. The calculation of $t\bar{t}$ differential cross sections, as shown in Eq. 2.31, involves a combination of partonic cross sections from perturbation theory and non-perturbative PDFs. The $t\bar{t}$ process can then be modelled using event generators, which are a combination matrix-element generators for the hard interaction and parton shower simulators for dressing the hard interaction with soft radiative emission. The event generator then fragments and hadronises the final state partons. At this level, the particle level, is what the unfolded cross section measurements will be compared to. In order to compare to the event yields at reconstruction level, a model of the CMS detector can be applied to the simulation.

Both in simulation and data, the final state deposits need to be reconstructed into particle candidates. This is done using the particle-flow algorithm, which categorises the energy and charge deposits into electrons, muons, photon, charged and neutral hadrons that are clustered into jets, and missing p_T . Further identification criteria are applied to the particle candidates to increase the purity of the selected objects.

Now with the event fully reconstructed and analysis objects defined, an event selection can be applied based on the final state of the single lepton $t\bar{t}$ decay channel. The reconstructed yields of $t\bar{t}$ events in data and simulation can then be measured with respect to the N_{jets} , H_T , S_T , p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ kinematic event variables. When algorithms are applied to both data and simulation, they very often yield different efficiencies. Scale factors are applied in simulation in order to match what is seen in data.

With the reconstructed yields calculated, the efficiency, acceptance and resolution effects from the CMS detector need to be undone, using a method called unfolding. Measurements of the normalised and absolute differential cross sections are performed to particle level in a visible phase space. Normalised cross section measurements have greater precision due to the cancellation of some systematics, whereas the absolute measurements contain all the informa-

tion available. They are presented to particle level in order to be useful for comparisons with respect to future models, where the simulation of the CMS experiment is the most time consuming part of model-making, and to measurements from other experiments. It allows for future models to be compared without wasting valuable time and computational resources on the simulation of the detector and avoids the influence of large theoretical uncertainties introduced by extrapolating the results to parton level (or a full phase space).

Finally, there is always an uncertainty associated with the cross section measurements. These come in the form of statistical and systematic uncertainties. The systematic uncertainties can come from many experimental and theoretical sources and are estimated by varying the simulation and/or the background predictions.

The differential cross sections measurements are presented here along with the results from χ^2 goodness-of-fit tests. Additional studies with respect to the theoretical uncertainties in the $t\bar{t}$ model and the effects of non-regularised unfolding are also shown.

9.1 Cross section measurements

Figure 73 shows the normalised differential cross sections with respect to the N_{jets} , H_T and S_T event variables and Fig. 74 with respect to the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ variables respectively. Similarly, Figs. 75 and 76 show the absolute cross section measurements. The cross section measurements are compared to four particle-level predictions, before the detector modelling, introduced in Sec. 4.6:

- POWHEG+PYTHIA
- POWHEG+HERWIG++
- MG5_aMC@NLO-NLO+PYTHIA
- MG5_aMC@NLO-LO+PYTHIA

In the lower panels, the ratio of the model cross section to the measured cross section is shown overlaid onto the grey statistical uncertainty band and the gold systematic band. By eye, good agreement can already be seen for the N_{jets} variable between the measured cross sections and the POWHEG+PYTHIA and MG5_aMC@NLO-NLO+PYTHIA models. The trend in the modelling of the p_T of the top quark can easily be seen for all the hadronically based variables and which is softer in data than simulation. Figures 77, 78, 79 and 80 show the normalised and absolute cross section measurements compared to three variations of the POWHEG+PYTHIA simulation. These are variations of the shower scales, variations of the h_{damp} matching parameter and a variation with the application of reweighting to the top quark p_T . In the latter case, the slope seen in the ratio is much less pronounced as expected.

While useful information can be inferred from these plots, absolute statements cannot be made regarding how well a simulation models the data. This is due to the correlations of the uncertainties between bins of the measurement. A χ^2 goodness-of-fit test needs to be performed between each model and the data.

9.2 Goodness-of-fit tests

The χ^2 goodness-of-fit test is defined by

$$\chi^2 = u_i \cdot \mathbf{V}_{ij}^{-1} \cdot u_j^T, \quad (9.1)$$

where u is a vector of the differences between the measured cross section and the model per bin $u_i = (x_i^{\text{meas}} - x_i^{\text{model}})$ and u^T is the transpose of this vector. The correlations between the bins are taken into account by the full covariance matrix, \mathbf{V} . The full covariance matrix is created by adding together all the covariance matrices for all statistical and systematic uncertainties. Each element in the covariance matrix is defined as the average of upward and downward uncertainties, U , according to:

$$V_{ij} = \frac{1}{2}(U_i^{\text{up}}U_j^{\text{up}} + U_i^{\text{down}}U_j^{\text{down}}). \quad (9.2)$$

In terms of systematic uncertainties which consist of only one variation, the elements of the covariance matrix are simply given by

$$V_{ij} = U_i U_j. \quad (9.3)$$

The total covariances for the normalised differential cross section measurements are given in Figs. 121 and 122 and for the absolute measurements in Figs. 123 and 124, all of which can be found in App. I

Each measurement has a number of degrees of freedom equal to its number of bins. This is reduced by one for the normalised cross section measurements because of the additional constraint from the normalisation. A p-value can be obtained for each χ^2/ndf . A second χ^2 test is performed between the measured cross sections and the POWHEG+PYTHIA model where the theoretical uncertainties in the model as well as in the unfolded data are taken into account. The correlations between the uncertainties in the prediction of the model and the measured cross section are included. All χ^2/ndf and p-values are shown per $t\bar{t}$ model in Tabs. 15 and 16 for the normalised and absolute cross section measurements respectively.

The predictions of the S_T , p_T^{miss} and p_T^ℓ kinematic event distributions by the POWHEG+PYTHIA model are consistent with the measured cross sections. The N_{jets} variable is well described by the POWHEG+PYTHIA generator with χ^2/ndf of 2.0/5 and 2.2/6 for the normalised and absolute cross section measurements respectively. The tune CUETP8M2T4 used in the POWHEG+PYTHIA model is derived using jet multiplicity measurements from 8 TeV, as seen in Fig. 11 and the good description measured here confirms the tune describes well the data for a larger data set at higher \sqrt{s} . Tensions are seen in the H_T , p_T^W and $|\eta^\ell|$ distributions. The results of the χ^2 tests with the inclusion of the theoretical uncertainties in the POWHEG+PYTHIA model show that any tensions observed are covered, with the exception of the $|\eta^\ell|$ distribution.

The POWHEG+HERWIG++ model is broadly consistent with the cross section measurements, particularly for the p_T^{miss} , p_T^ℓ and p_T^W event variables, however large tensions are seen for the N_{jets} distributions. Tension is also seen for the $|\eta^\ell|$ variable. The MG5_aMC@NLO-NLO+PYTHIA model is consistent with all the kinematic event variables measured, with the exception of $|\eta^\ell|$. If the theoretical uncertainties are applied it is expected that it would cover some of the tensions seen in both the POWHEG+HERWIG++ and MG5_aMC@NLO-NLO+PYTHIA models. Without considering the theoretical uncertainties in the MG5_aMC@NLO-LO+PYTHIA model, there is no consistency with measured data for any variable.

No model is consistent for all the cross section measurements without taking the theoretical uncertainties in the model into account. The NLO models are more consistent with the data than the LO model. When taking into account model theoretical uncertainties, the tensions are reduced. The $|\eta^\ell|$ event variable causes tensions in all models studied. Tables 20 and 20, as well as Fig. 125, all shown in App. J, show that the tension with respect to $|\eta^\ell|$ stems from the μ +jets decay channel. The χ^2/ndf in the e+jets decay channel is never worse than 6.5/7 for the POWHEG+PYTHIA model. The tension in the μ +jets decay channel is most likely caused

by an effect on the shape of the distribution, particularly in the final bin. One possibility is that it could come from the scale factors that have been applied.

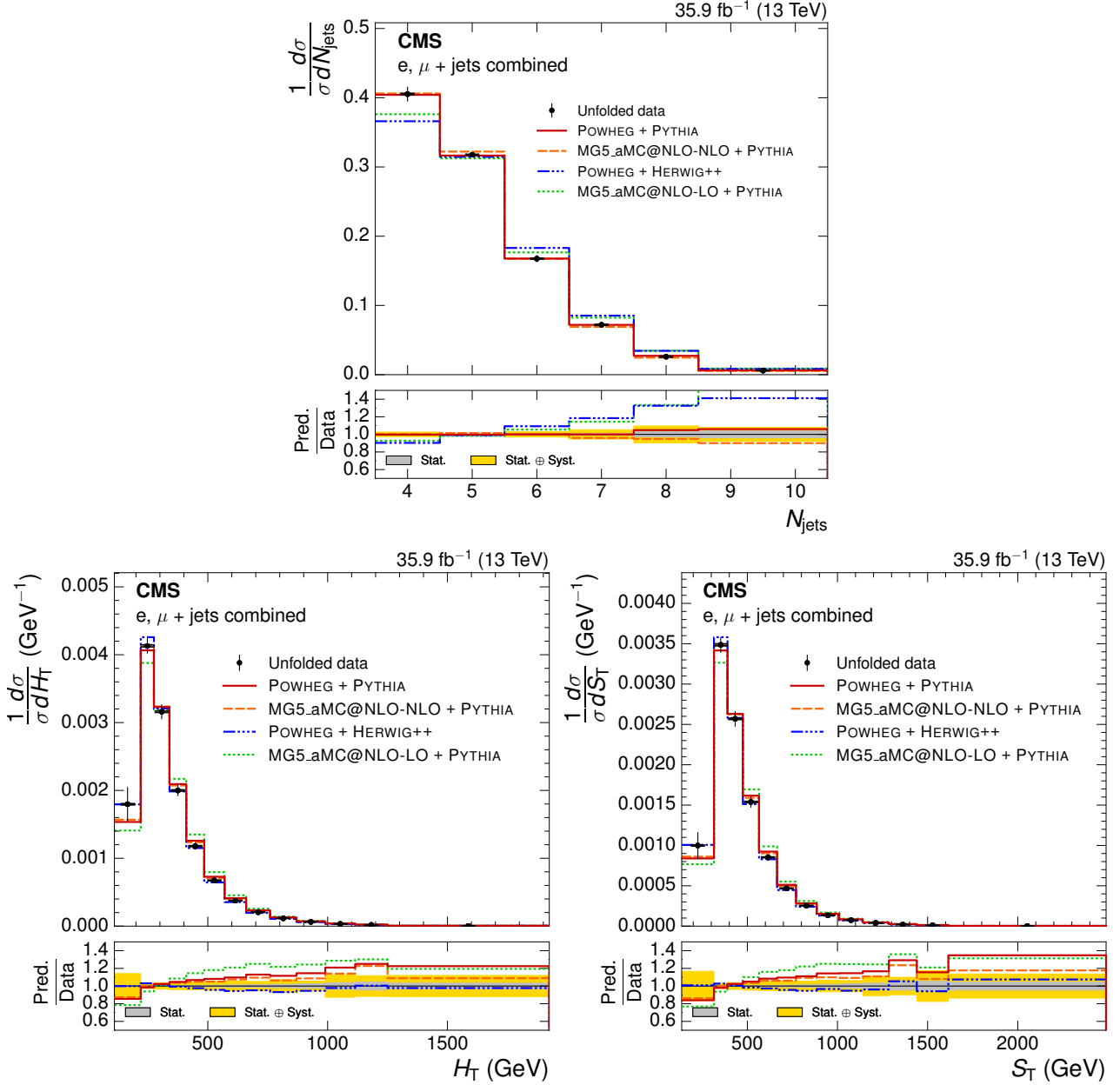


Figure 73: The normalised cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables. They are compared against different $t\bar{t}$ simulations. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

9.3 Effects of regularisation

A study on the effects of the regularisation in the unfolding on the χ^2/ndf calculation is performed. A lack of regularisation can lead to non-physical effects present in the unfolded distributions, however too much regularisation and the unfolded data is biased towards the input simulation.

Tables 17 and 18 show the χ^2/ndf produced when regularisation is not used. Typically only small changes are seen between the regularised and unregularised measurements. The

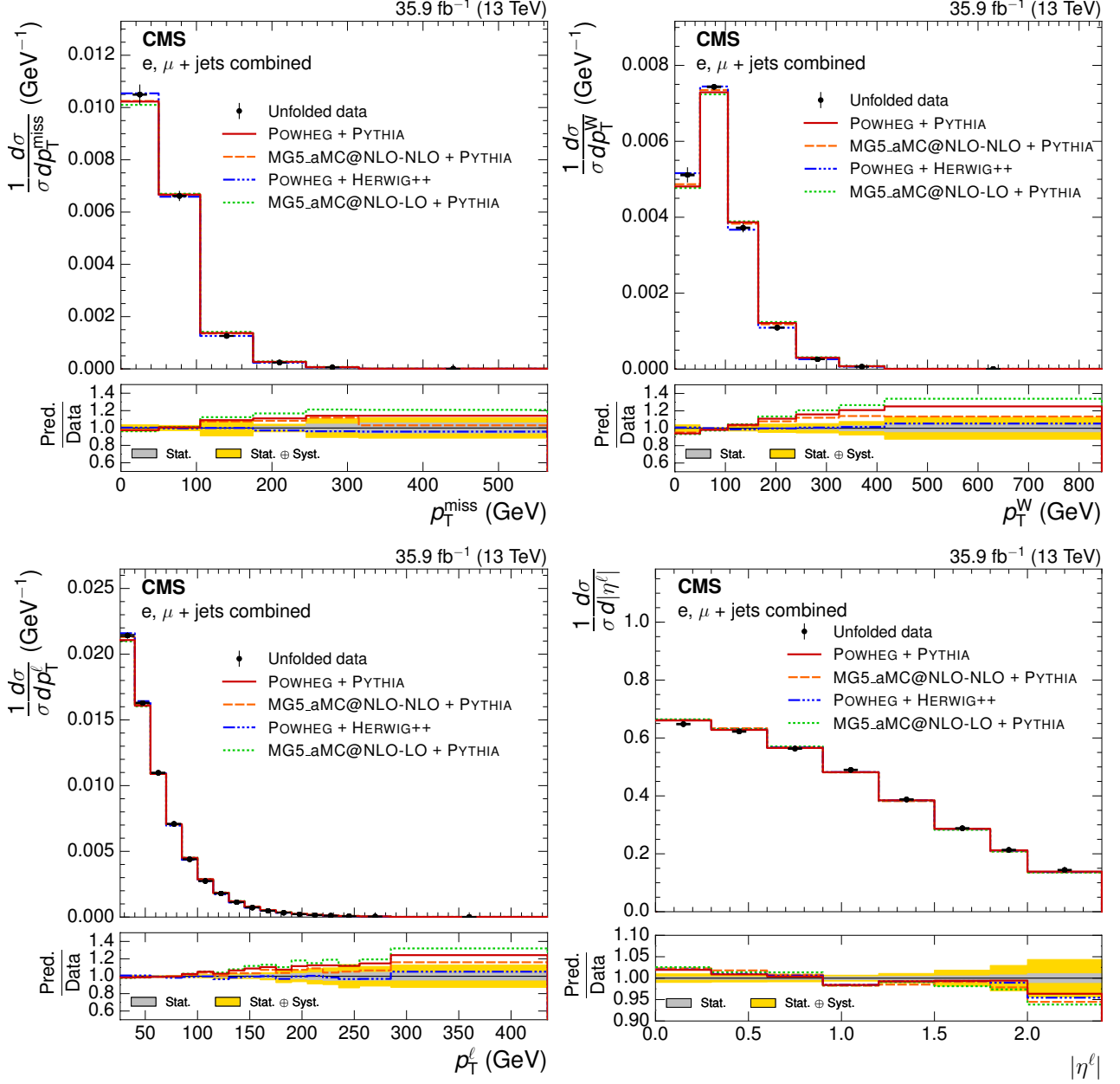


Figure 74: The normalised cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ kinematic event variables. They are compared against different $t\bar{t}$ simulations. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

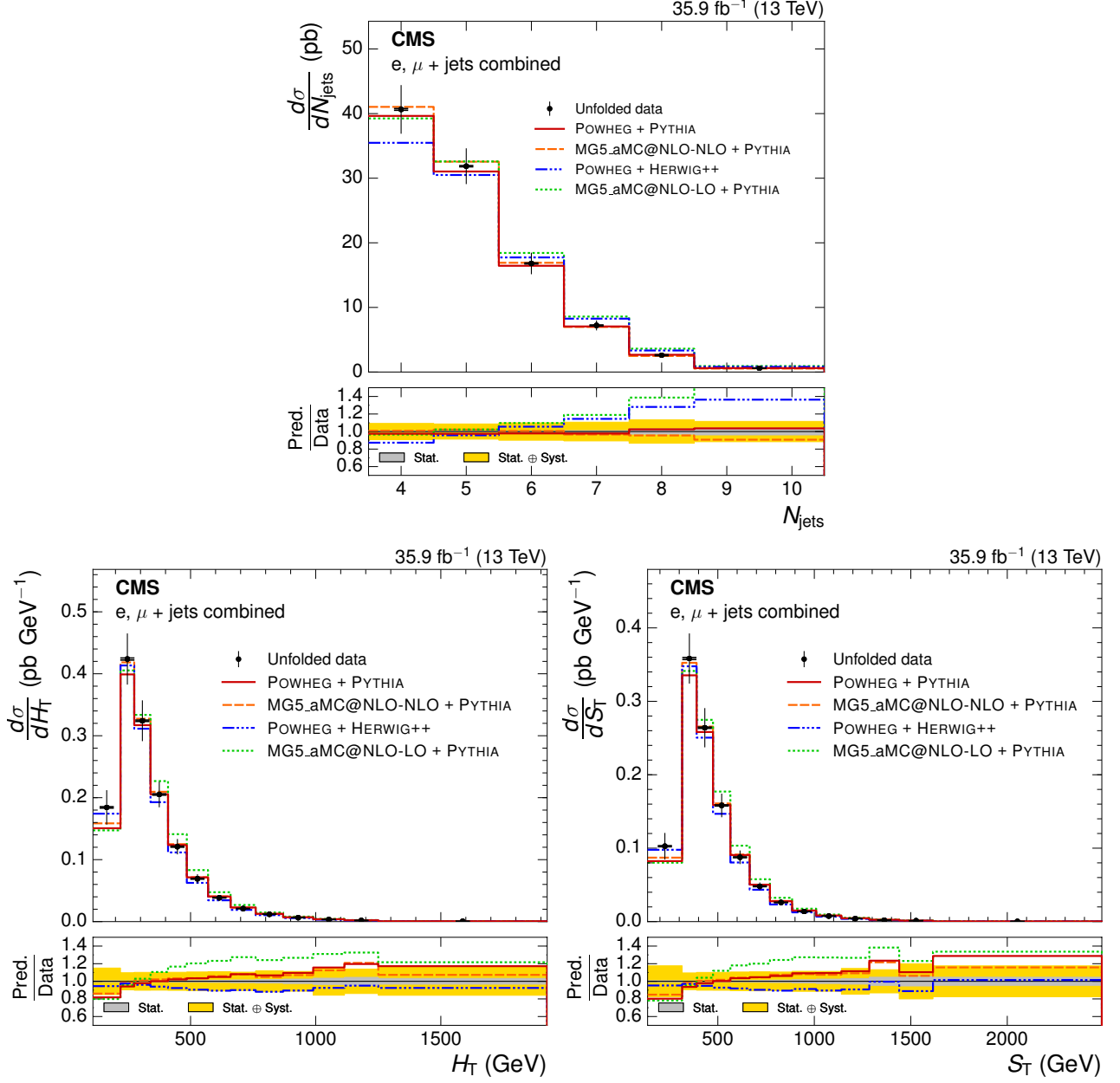


Figure 75: The absolute cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables. They are compared against different $t\bar{t}$ simulations. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

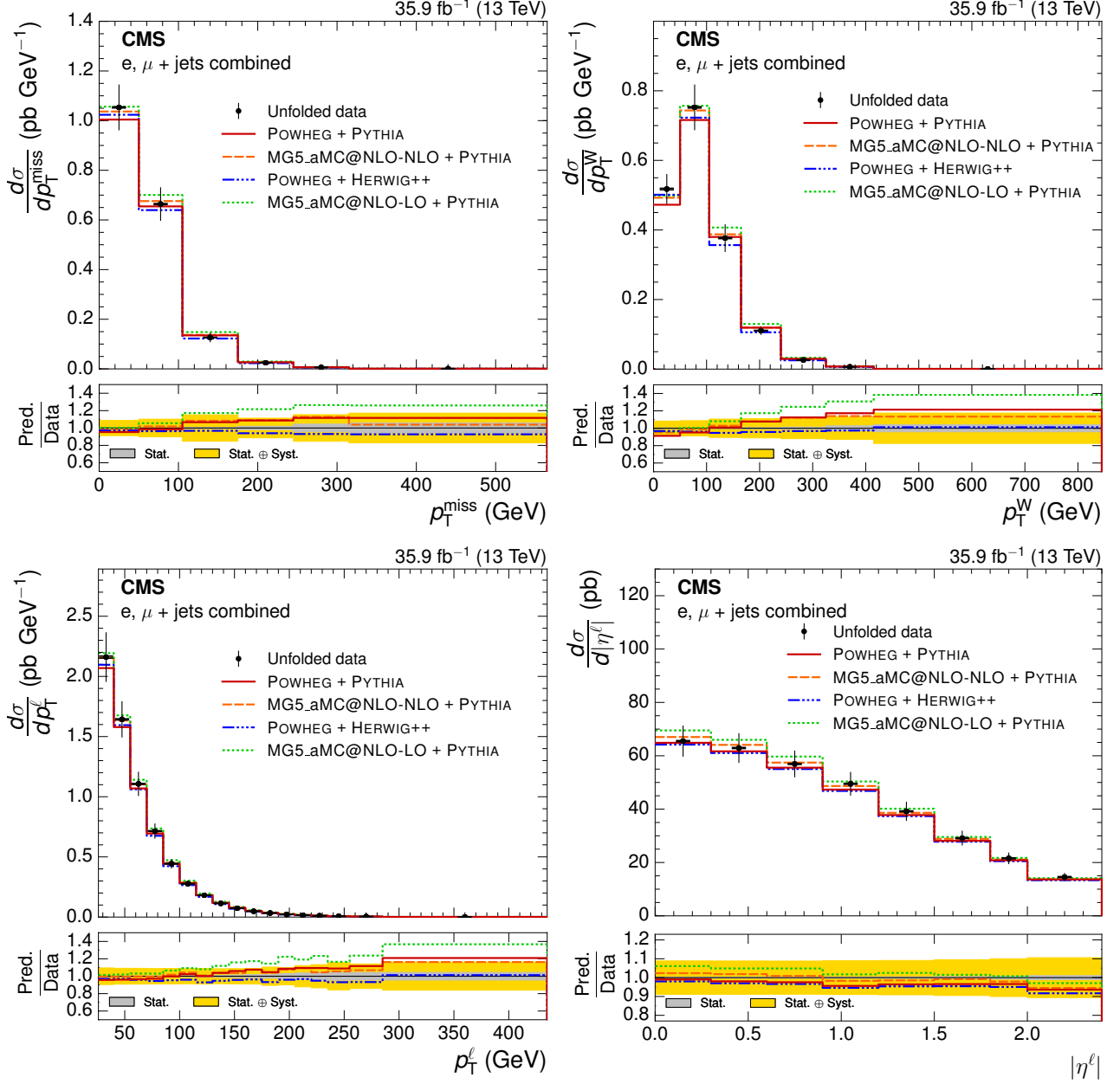


Figure 76: The absolute cross section measurements as a function of the p_T^{miss} , p_T^{W} , p_T^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against different $t\bar{t}$ simulations. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

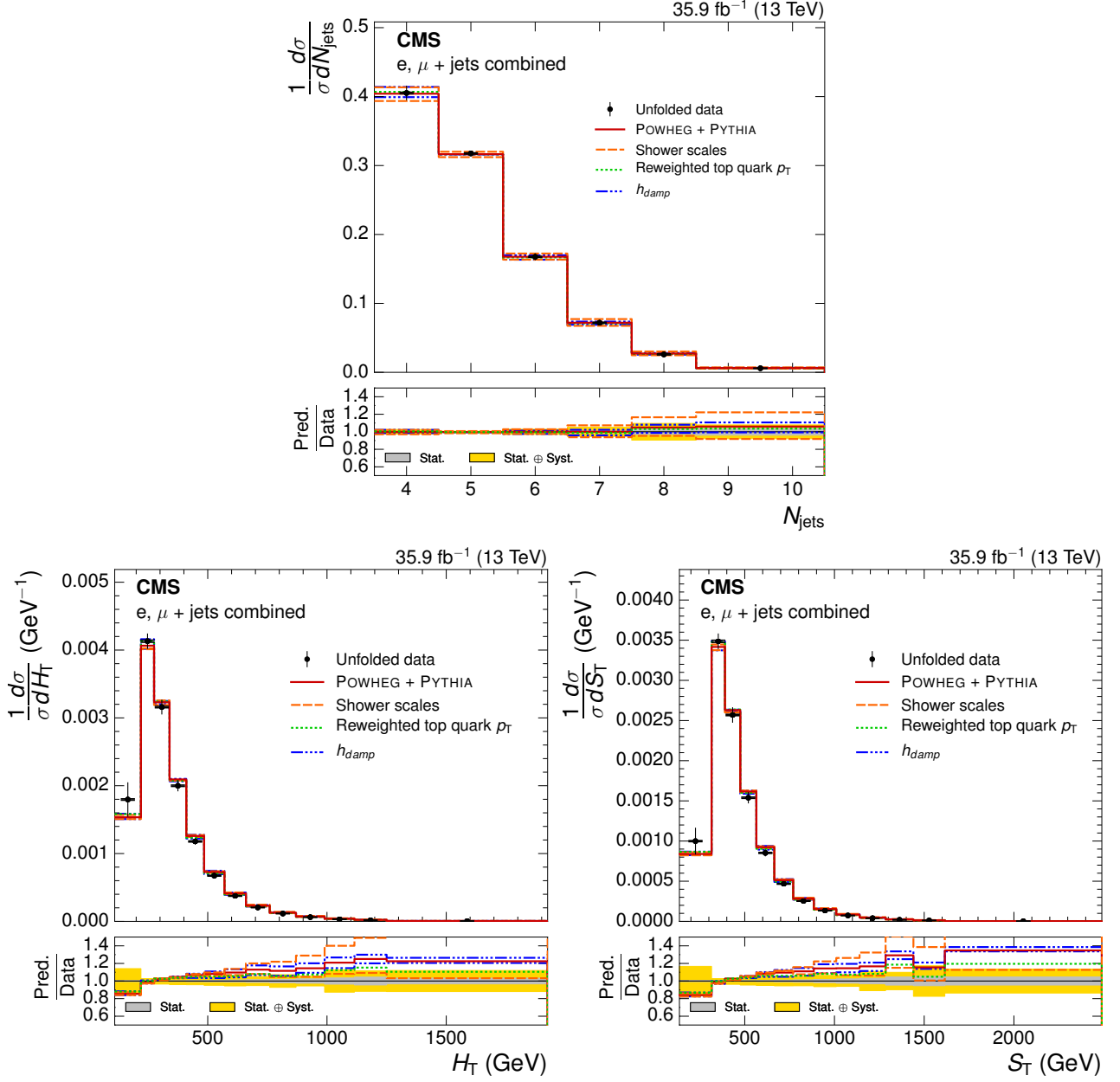


Figure 77: The normalised cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

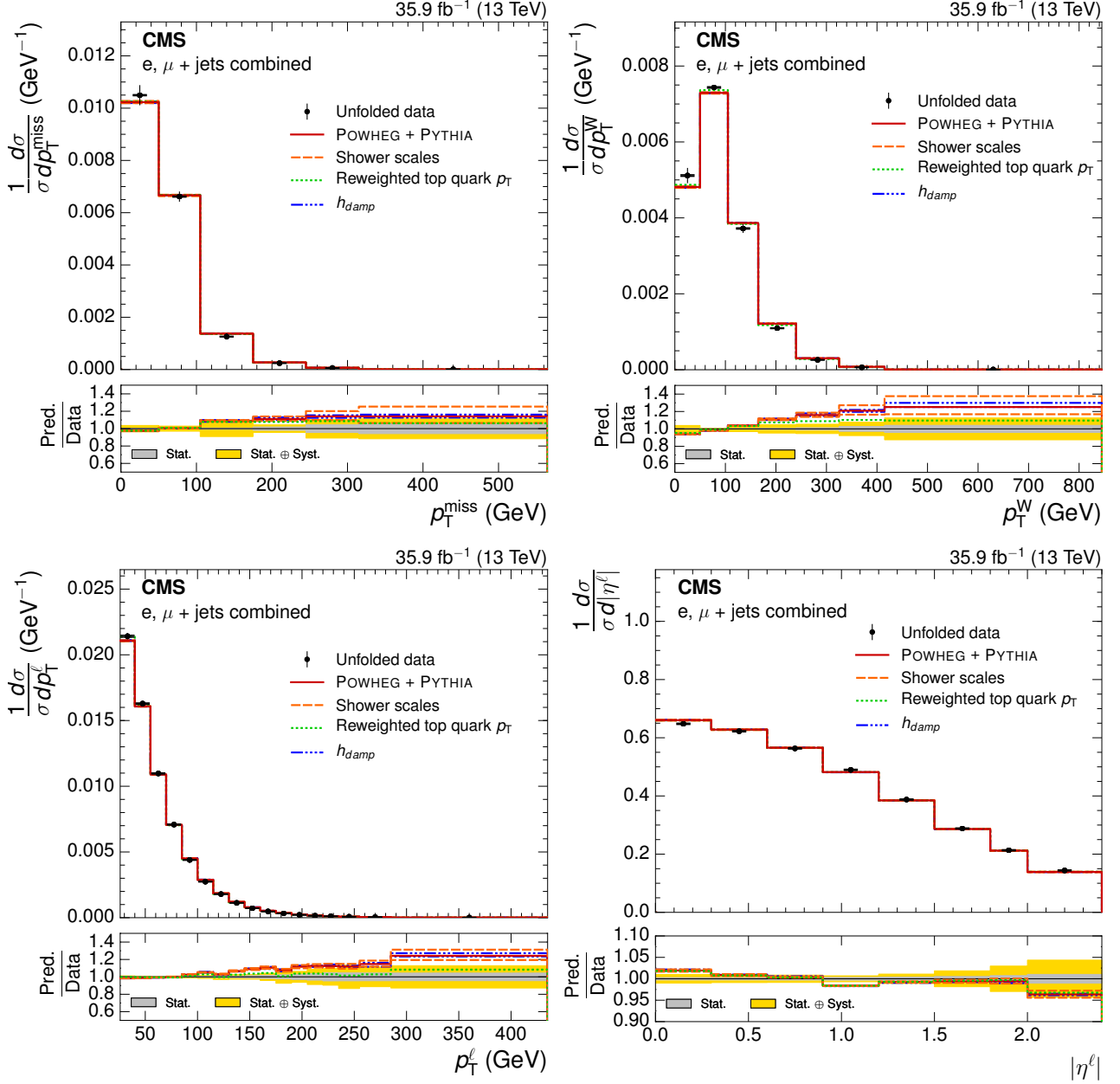


Figure 78: The normalised cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

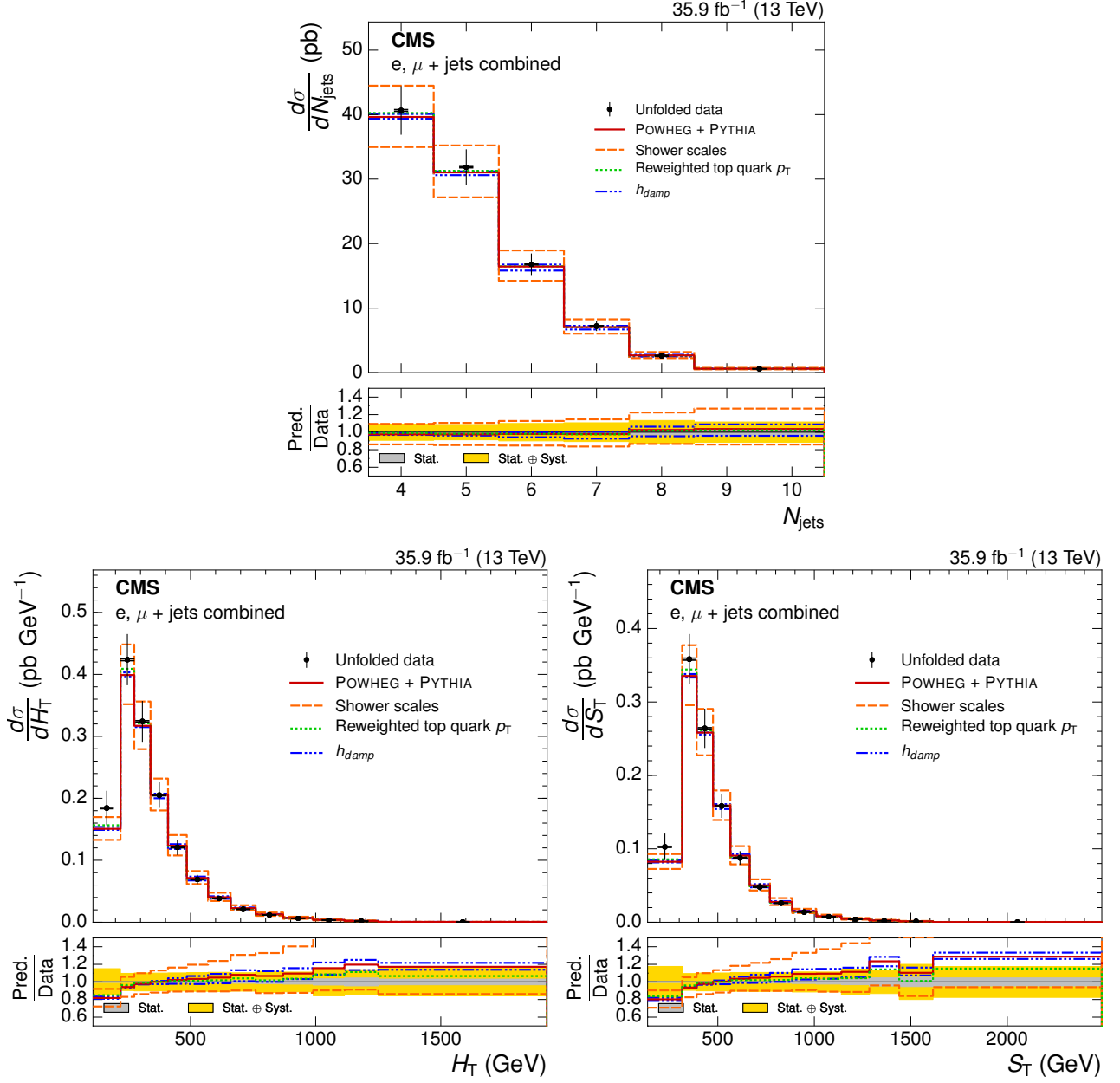


Figure 79: The absolute cross section measurements as a function of the N_{jets} , H_T and S_T kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

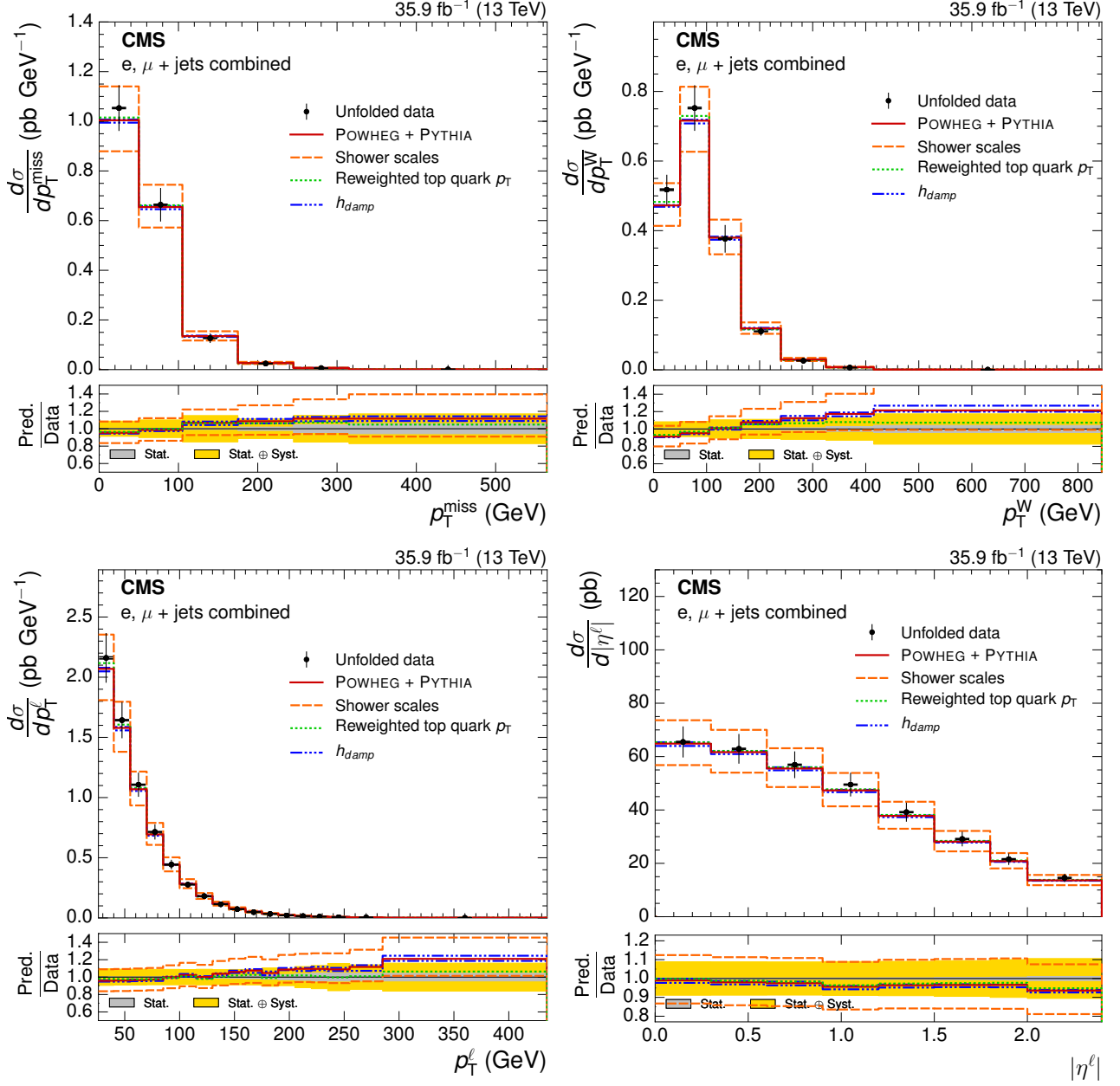


Figure 80: The absolute cross section measurements as a function of the p_T^{miss} , p_T^W , p_T^l and $|\eta^l|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The vertical bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Table 15: Results of a goodness-of-fit test between the normalised cross sections in data and several models, with values given as $\chi^2/\text{number of degrees of freedom (ndf)}$.

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	2.0 / 5	0.85	1.5 / 5	0.91
H_{T}	26 / 12	< 0.01	4.8 / 12	0.97
S_{T}	22 / 12	0.04	4.2 / 12	0.98
$p_{\text{T}}^{\text{miss}}$	11 / 5	0.06	2.9 / 5	0.72
p_{T}^{W}	16 / 6	0.01	2.5 / 6	0.87
p_{T}^{ℓ}	24 / 16	0.09	14 / 16	0.63
$ \eta^{\ell} $	19 / 7	< 0.01	15 / 7	0.04

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	38 / 5	< 0.01	9.5 / 5	0.09	78 / 5	< 0.01
H_{T}	23 / 12	0.03	11 / 12	0.52	160 / 12	< 0.01
S_{T}	21 / 12	0.04	11 / 12	0.57	110 / 12	< 0.01
$p_{\text{T}}^{\text{miss}}$	1.3 / 5	0.93	5.9 / 5	0.31	23 / 5	< 0.01
p_{T}^{W}	0.81 / 6	0.99	8.9 / 6	0.18	30 / 6	< 0.01
p_{T}^{ℓ}	11 / 16	0.82	16 / 16	0.44	37 / 16	< 0.01
$ \eta^{\ell} $	19 / 7	< 0.01	24 / 7	< 0.01	30 / 7	< 0.01

Table 16: Results of a goodness-of-fit test between the absolute cross sections in data and several models, with values given as $\chi^2/\text{number of degrees of freedom (ndf)}$.

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	2.2 / 6	0.90	1.7 / 6	0.95
H_{T}	23 / 13	0.05	4.3 / 13	0.99
S_{T}	19 / 13	0.11	4.7 / 13	0.98
$p_{\text{T}}^{\text{miss}}$	13 / 6	0.05	3.1 / 6	0.80
p_{T}^{W}	17 / 7	0.02	2.7 / 7	0.91
p_{T}^{ℓ}	20 / 17	0.28	14 / 17	0.68
$ \eta^{\ell} $	16 / 8	0.04	15 / 8	0.06

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	39 / 6	< 0.01	12 / 6	0.07	93 / 6	< 0.01
H_{T}	21 / 13	0.07	10 / 13	0.66	150 / 13	< 0.01
S_{T}	18 / 13	0.17	9.3 / 13	0.75	110 / 13	< 0.01
$p_{\text{T}}^{\text{miss}}$	1.5 / 6	0.96	6.6 / 6	0.36	26 / 6	< 0.01
p_{T}^{W}	0.90 / 7	1.00	9.2 / 7	0.24	33 / 7	< 0.01
p_{T}^{ℓ}	11 / 17	0.87	15 / 17	0.58	36 / 17	< 0.01
$ \eta^{\ell} $	17 / 8	0.04	23 / 8	< 0.01	31 / 8	< 0.01

largest changes are, as expected, linked to event variables which have a low experimental resolution. One of the largest changes in the χ^2/ndf is seen in the $p_{\text{T}}^{\text{miss}}$ distribution of the POWHEG+PYTHIA simulation with model theoretical uncertainties where it decreases from 2.9/5 to 2.1/5. Another is for the H_{T} distribution in the MG5_aMC@NLO-NLO+PYTHIA model where it increases from 11/12 to 12/12. The χ^2/ndf does not change significantly for most other event variables and not at all for p_{T}^{ℓ} and $|\eta^{\ell}|$. This implies that regularisation is not needed in the case of variables with a high experimental resolution.

Table 17: Results of a goodness-of-fit test between the unregularised normalised cross sections in data and several models, with values given as $\chi^2/\text{number of degrees of freedom (ndf)}$.

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	1.9 / 5	0.86	1.2 / 5	0.95
H_{T}	25 / 12	0.01	5.6 / 12	0.94
S_{T}	22 / 12	0.04	5.6 / 12	0.94
$p_{\text{T}}^{\text{miss}}$	9.3 / 5	0.10	2.1 / 5	0.84
p_{T}^{W}	16 / 6	0.01	2.7 / 6	0.84
p_{T}^{ℓ}	24 / 16	0.09	13 / 16	0.64
$ \eta^{\ell} $	19 / 7	< 0.01	15 / 7	0.04

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	42 / 5	< 0.01	7.6 / 5	0.18	81 / 5	< 0.01
H_{T}	21 / 12	0.05	12 / 12	0.44	140 / 12	< 0.01
S_{T}	22 / 12	0.04	11 / 12	0.54	98 / 12	< 0.01
$p_{\text{T}}^{\text{miss}}$	0.82 / 5	0.98	5.2 / 5	0.39	21 / 5	< 0.01
p_{T}^{W}	1.7 / 6	0.94	9.4 / 6	0.15	29 / 6	< 0.01
p_{T}^{ℓ}	11 / 16	0.83	16 / 16	0.44	37 / 16	< 0.01
$ \eta^{\ell} $	19 / 7	< 0.01	24 / 7	< 0.01	30 / 7	< 0.01

Table 18: Results of a goodness-of-fit test between the unregularised absolute cross sections in data and several models, with values given as $\chi^2/\text{number of degrees of freedom (ndf)}$.

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	1.7 / 6	0.94	1.5 / 6	0.96
H_{T}	21 / 13	0.07	4.8 / 13	0.98
S_{T}	18 / 13	0.15	6.0 / 13	0.94
$p_{\text{T}}^{\text{miss}}$	11 / 6	0.10	2.4 / 6	0.88
p_{T}^{W}	16 / 7	0.02	2.9 / 7	0.89
p_{T}^{ℓ}	20 / 17	0.28	14 / 17	0.68
$ \eta^{\ell} $	16 / 8	0.04	15 / 8	0.06

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	38 / 6	< 0.01	7.2 / 6	0.31	70 / 6	< 0.01
H_{T}	20 / 13	0.09	11 / 13	0.64	140 / 13	< 0.01
S_{T}	19 / 13	0.14	9.8 / 13	0.71	100 / 13	< 0.01
$p_{\text{T}}^{\text{miss}}$	0.86 / 6	0.99	5.7 / 6	0.45	25 / 6	< 0.01
p_{T}^{W}	1.9 / 7	0.96	9.7 / 7	0.21	31 / 7	< 0.01
p_{T}^{ℓ}	10 / 17	0.88	15 / 17	0.58	36 / 17	< 0.01
$ \eta^{\ell} $	17 / 8	0.04	23 / 8	< 0.01	31 / 8	< 0.01

CHAPTER 10

Conclusions, prospects and outlook

Both the normalised and absolute cross sections are measured differentially as a function of several kinematic event variables at particle level in a visible phase space. The results are compared to the state-of-the-art $t\bar{t}$ production models: POWHEG+PYTHIA, POWHEG+HERWIG++, MG5_aMC@NLO-LO+PYTHIA and MG5_aMC@NLO-NLO+PYTHIA. The goodness-of-fit tests performed between the simulations and measured cross section find that the POWHEG+PYTHIA model is generally consistent with the data, with any residual differences covered by theoretical uncertainties within the model. The POWHEG+HERWIG++ and MG5_aMC@NLO-NLO+PYTHIA are also consistent with the data for most of the kinematic event variables, whereas the MG5_aMC@NLO-LO+PYTHIA model is found not to accurately describe any variable.

It is widely expected for these measurements to be used in the tuning of future $t\bar{t}$ generators and therefore the measurements presented here have been implemented in RIVET and are available to the wider community. In addition, these measurements can be used for constraints on the PDF of the proton, as an accurate background estimate for beyond the SM or rare SM searches and as an input to phenomenology studies.

At present, the results are dominated by systematic uncertainties. The precision of these measurements can only be increased by lowering the so-called *systematic wall*. This can be achieved from an improved understanding of the theoretical uncertainties, e.g. using a model with a more accurate description of the parton shower, or from finding better ways to treat experimental uncertainties.

It makes sense then to go beyond the bounds of the current analysis into regions that are no longer dominated by the systematic uncertainty. This could be in the form of double or even triple differential cross sections using the data recorded in 2016. Going further and considering the total luminosity delivered by the LHC during Run 2, $> 150 \text{ fb}^{-1}$, it would make these measurements very precise, and very desirable. The ability to compare a model to two or three different variables simultaneously is an ideal prospect of this analysis for the $t\bar{t}$ generator modelling and tuning community. Important distributions, for example the $(p_T^{\text{top}}, |\eta^{\text{top}}|)$ and $(p_T^{\text{top}}, M_{t\bar{t}})$ distributions have already been studied [90]. These allow for comparisons to the full kinematic properties of the top quark as well as information on the inconsistencies seen between the top p_T spectrum and that of the mass of the $t\bar{t}$ system. The future of this work could provide useful additional distributions with respect to the number of additional jets or

missing p_T of the system.

Another option is to go beyond the kinematic event variables used in this analysis. They could include variables such as the angular separation between the two most likely b jets, the angular separation between the lepton and closest b jet or the p_T of additional jets. New and novel event variables, for example the n-Jettiness of the event [179], could also be measured. Of course, all the event variables currently measured can also be extrapolated to parton level in a full phase space, although perhaps more interesting and useful would be to extend them into the boosted regime, i.e. the regime where the top decay products become highly collimated due to being produced at high energy.

Finally, an option is to reinterpret the analysis as a part of an effective field theory (EFT) [180]. An EFT assumes that any new physics is decoupled from the SM and present at an energy scale inaccessible to current accelerator technology. As such, the SM is valid only up to this energy scale, and any new physics would then be introduced by deviations seen in the tails of the distributions caused by the interference between the SM and EFT. The RIVET toolkit is a powerful tool in the reinterpretation of analyses.

The following sections introduce some preliminary work, based on these future possibilities for the measurements presented in this analysis.

10.1 Double differential cross sections

A few preliminary double differential cross section measurements have been performed with respect to the (S_T, N_{jets}) , $(p_T^\ell, |\eta^\ell|)$, (S_T, p_T^{miss}) and $(p_T^{\text{miss}}, N_{\text{jets}})$ event variable pairs. For each double differential cross section measurement, a set of plots of the data-simulation comparison in the e+jets and μ +jets channels, the migration matrices for unfolding, and the normalised and absolute differential cross section measurements are shown. Figures 81, 82 and 83 show the relevant distributions with respect to (S_T, N_{jets}) and Fig. 84, 85 and 86 with respect to $(p_T^\ell, |\eta^\ell|)$. Further distributions for (S_T, p_T^{miss}) and $(p_T^{\text{miss}}, N_{\text{jets}})$ are shown in Figs. 126, 127, 128, 129, 130 and 131 in App. K.

The binning scheme for each variable has been chosen arbitrarily and the yield from multijet QCD measured from simulation. The $t\bar{t}$ yield is extracted using the background subtraction method, as for the single differential cross section measurements. The migration matrices used in the unfolding of the $t\bar{t}$ yields are created using the POWHEG+PYTHIA sample. The $t\bar{t}$ yields are unfolded without regularisation and so an unknown amount of bias may be present. The bias is thought to be negligible in the case of the $(p_T^\ell, |\eta^\ell|)$ double differential measurements, because both variables in the single differential measurements required only a minimal amount of regularisation. The unfolded $t\bar{t}$ yields are used to calculate the normalised and absolute cross sections to particle level in the same visible phase space as for the single differential measurements. A comparison to the POWHEG+PYTHIA $t\bar{t}$ production model is also shown. Only statistical uncertainties have been propagated through the measurement, at no point have any systematic uncertainties been evaluated or unfolding checks performed. The next steps would be to perform a full analysis, for which the most important aspects would be the quality tests on the unfolding procedure and the evaluation of the systematic uncertainties.

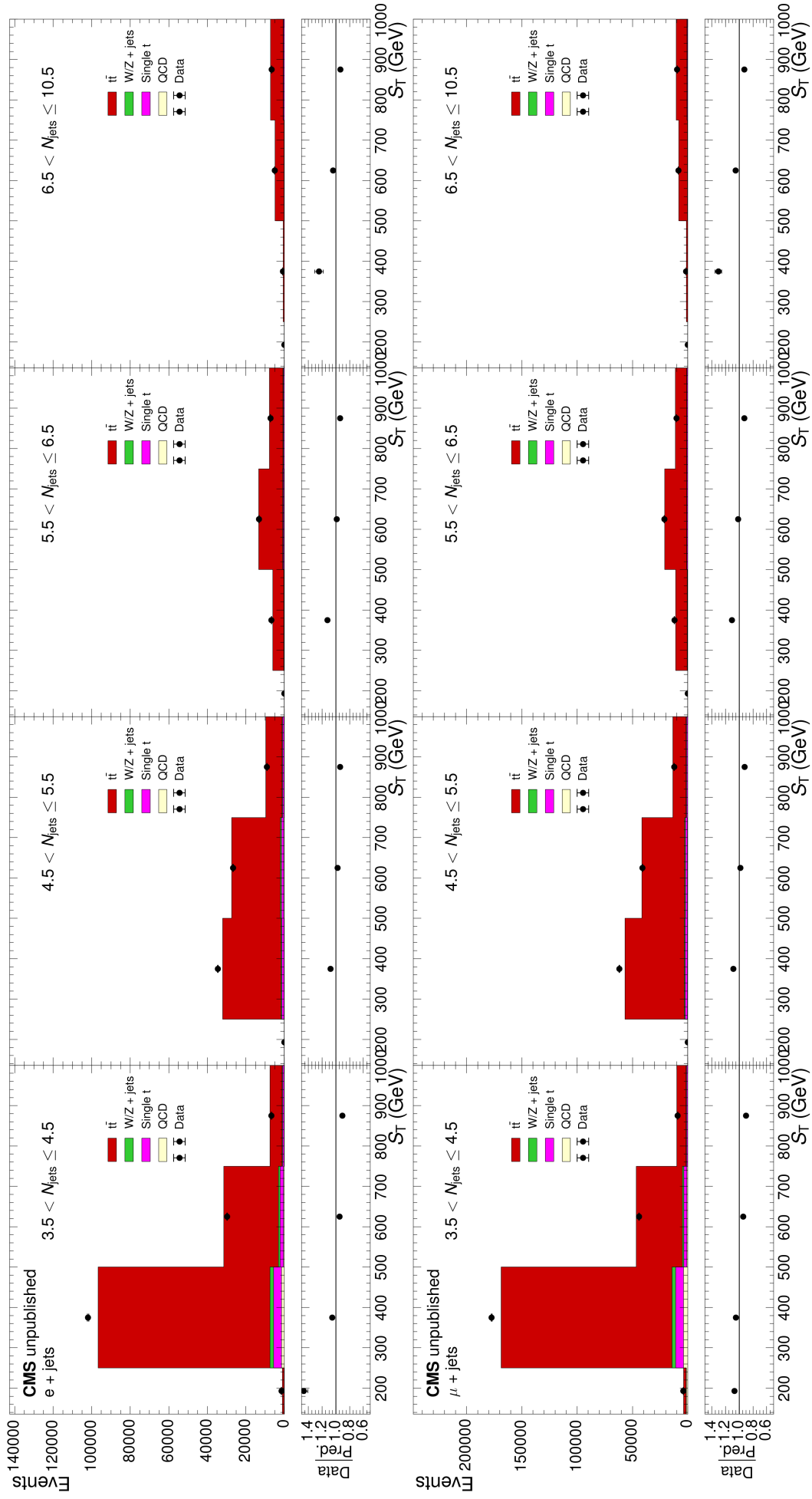


Figure 81: The $t\bar{t}$ yield after full event selection in the $e+\text{jets}$ (upper) and $\mu+\text{jets}$ (lower) channels as a function of the S_T and N_{jets} events variables. Only statistical uncertainties are shown.

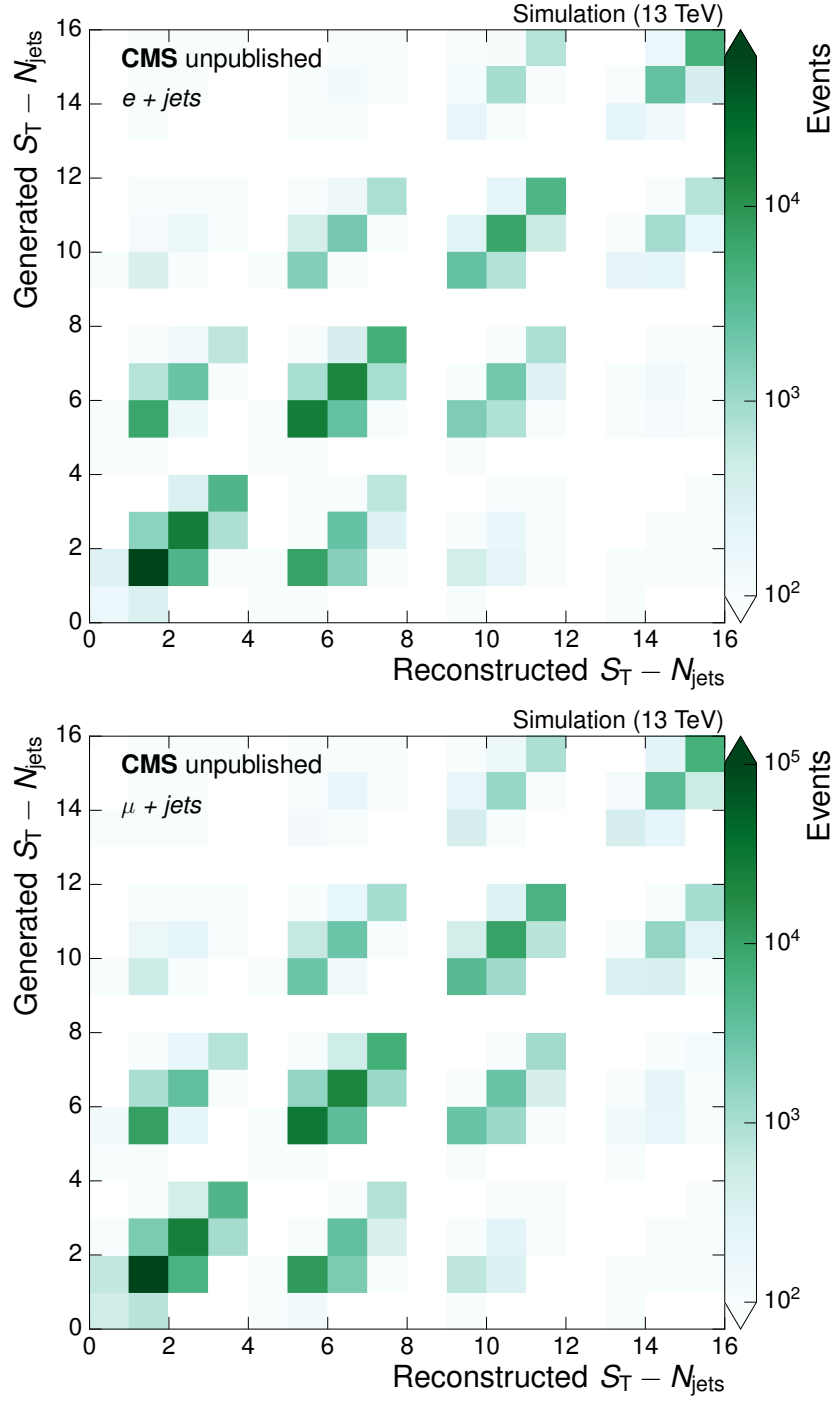


Figure 82: The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the S_T and N_{jets} events variables, in the $e + \text{jets}$ (upper) and $\mu + \text{jets}$ (lower) channels.

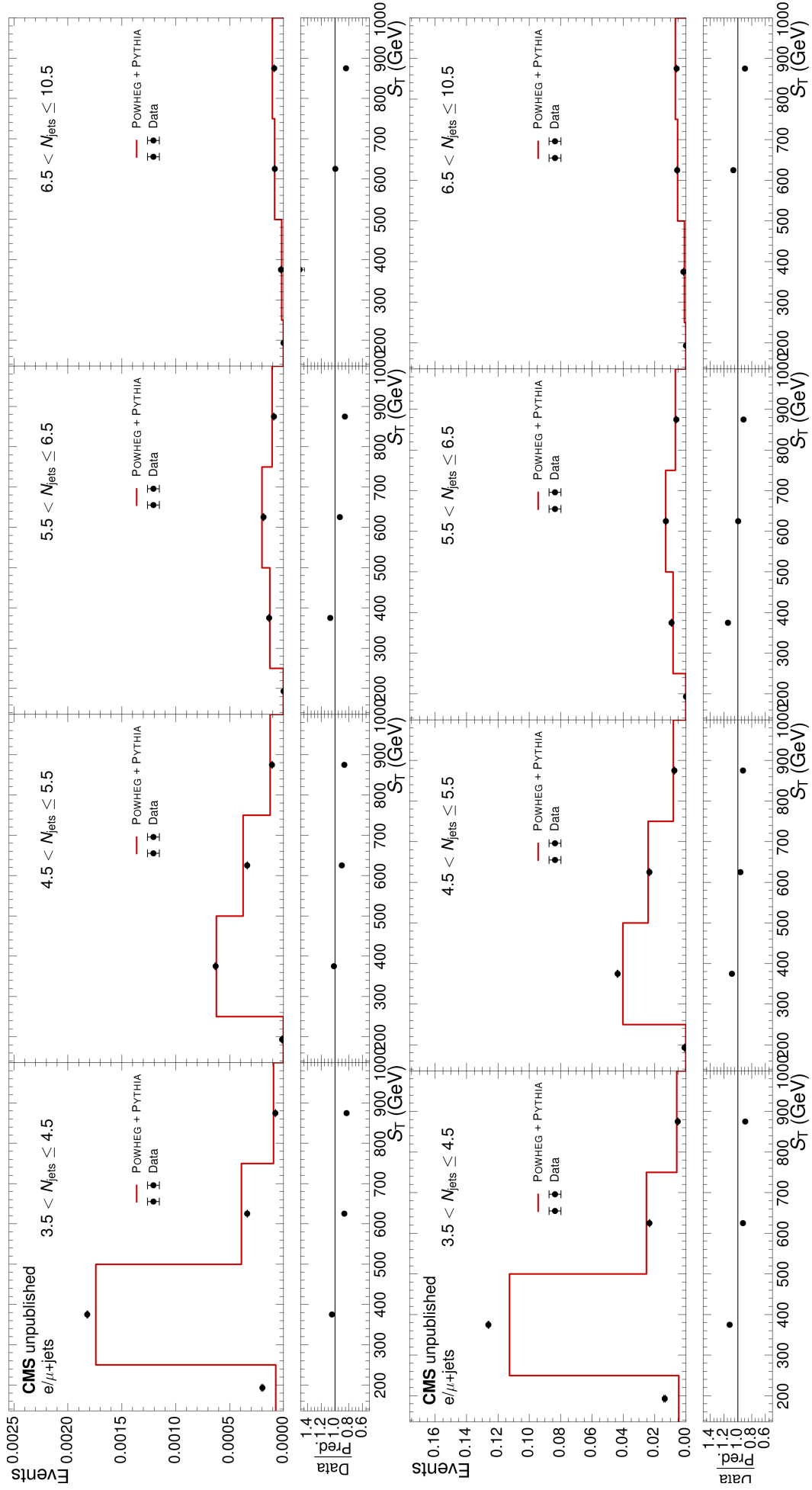


Figure 83: Normalised (upper) and absolute (lower) double differential $t\bar{t}$ production cross section measurements with respect to the S_T and N_{jets} event variables in comparison to the POWHEG+PYTHIA model. Only statistical uncertainties are shown.

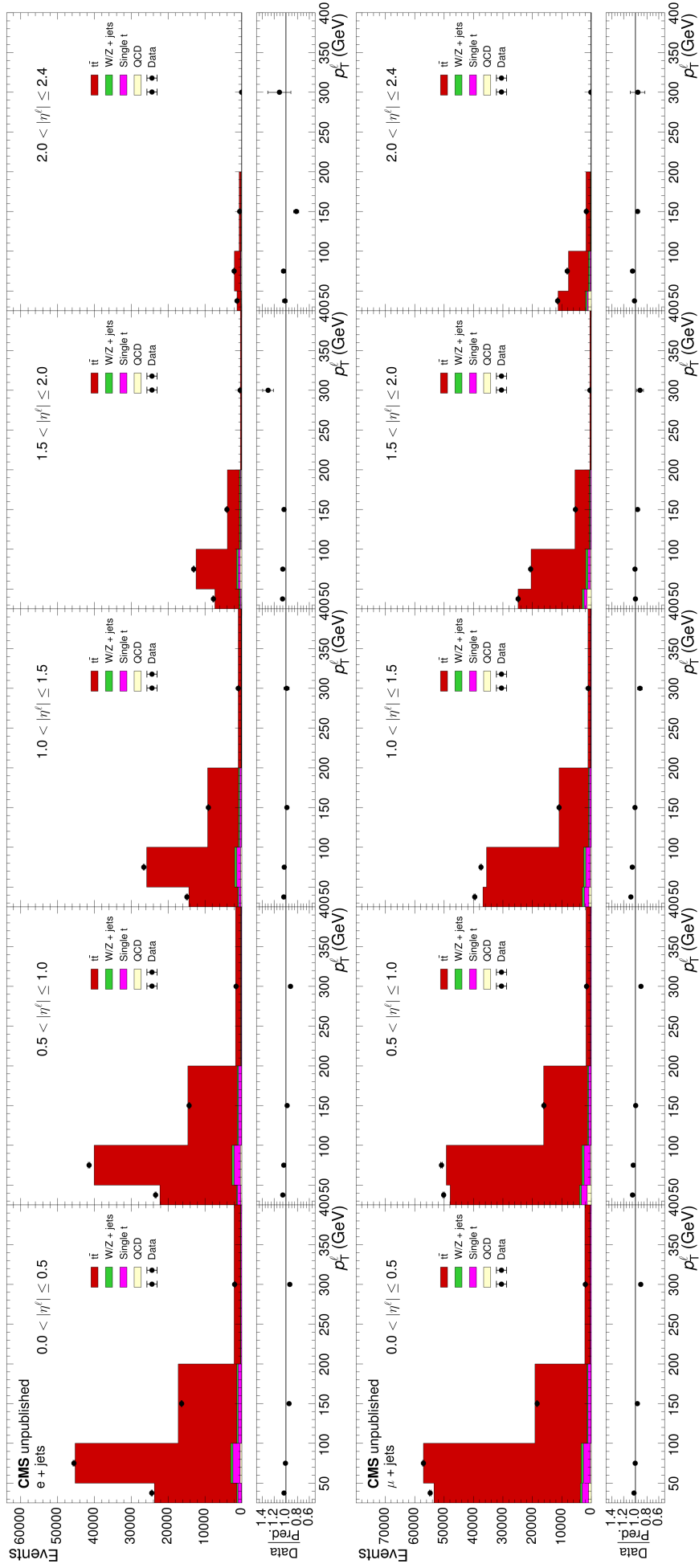


Figure 84: The $t\bar{t}$ yield after full event selection in the e +jets (upper) and μ +jets (lower) channels as a function of the p_T^ℓ and $|\eta^\ell|$ variables. Only statistical uncertainties are shown.

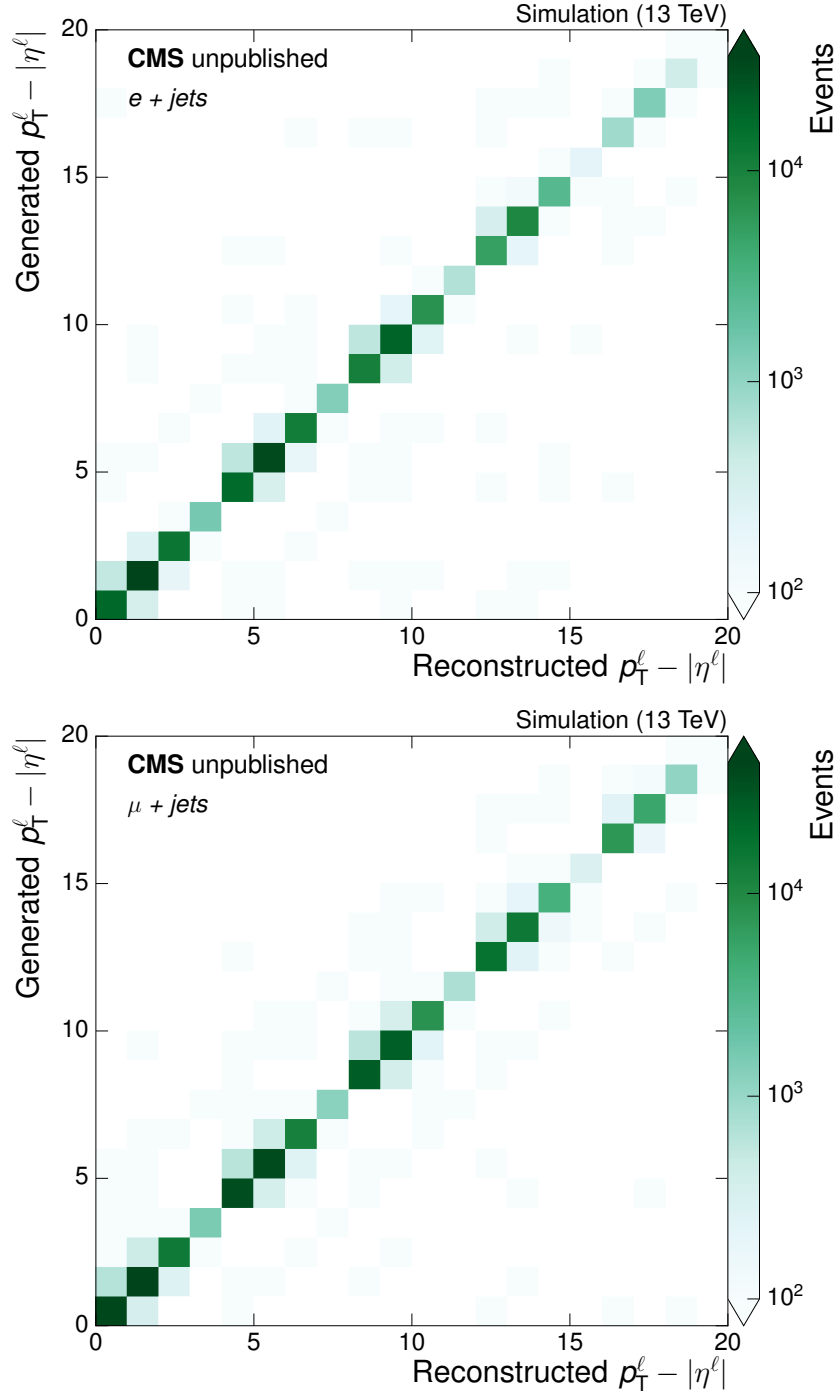


Figure 85: The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the p_T^ℓ and $|\eta^\ell|$ events variables, in the $e + \text{jets}$ (upper) and $\mu + \text{jets}$ (lower) channels.

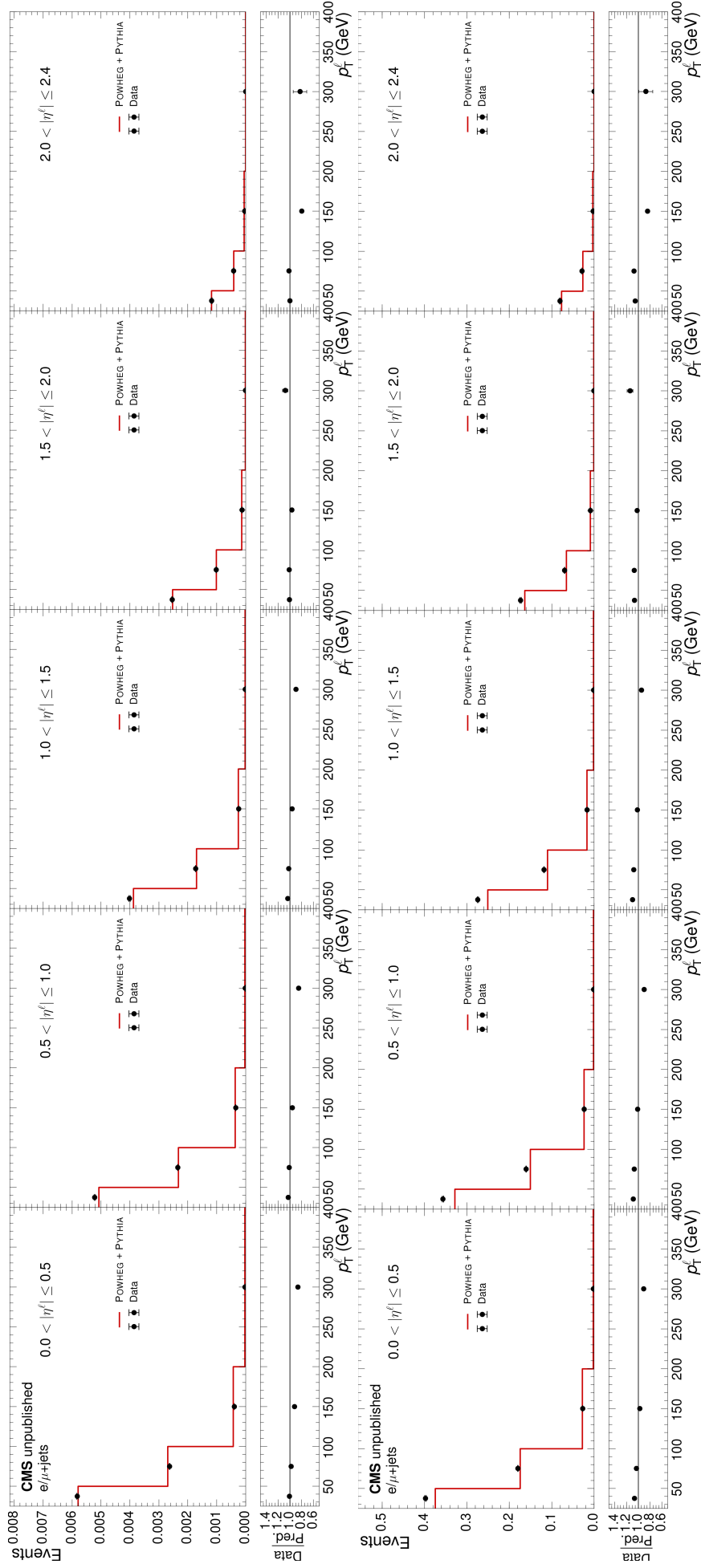


Figure 86: Normalised (upper) and absolute (lower) double differential $t\bar{t}$ production cross section measurements with respect to the p_T^l and $|\eta^l|$ event variables in comparison to the POWHEG+PYTHIA model. Only statistical uncertainties are shown.

10.2 Additional event variables

A set of possible alternative future variables are presented as simulation-data control plots, shown in Fig. 87. Only the statistical uncertainties are shown. The first is the mass of the signal lepton and its closest b quark, $M_{b\ell}$. This is defined by adding the respective four-momenta of the particles together and calculating

$$M_{b\ell} = \sqrt{E_{b\ell}^2 - \vec{p}_{b\ell}^2}, \quad (10.1)$$

for the system. As well as the mass of the $b\ell$ system, also shown is the angle between them, $\theta_{b\ell}$. The mass of the system of the leading three jets, M_3 , is measured in a similar way to $M_{b\ell}$. Finally, the scalar sum of the p_T of the 3rd and 4th leading jets, p_T^{j3j4} is shown along with the p_T of any additional jets in the event. Measurements with respect to these event variables provide additional information on the kinematic properties of the top quark system, for example acting as proxies for the masses of the top quark and W boson, and adding additional sensitivity to underlying event and ISR/FSR. These new variables can also be used as a part of a suite of double or triple differential cross section measurements.

10.3 Effective field theory

The SM is largely assumed to be a valid theory up until an energy scale at which new physics is introduced, Λ_{NP} . At this point any new field theory must satisfy three conditions, namely: the SM gauge group $\text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y$ should be contained within the new gauge group, all the degrees of freedom of the SM should still be present, and at low energies it should reduce into the SM. This reduction is believed to originate from the decoupling of the massive new particles at Λ_{NP} .

The new physics can be added to the SM Lagrangian by adding higher-dimensional operators, \mathcal{O}_i , constructed only out of SM fields, which maintains gauge invariance, and their associated coupling strengths, C_i , known as Wilson coefficients.

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_{d=5}^{\infty} \frac{1}{\Lambda_{\text{NP}}^{d-4}} \sum_i C_i^d \mathcal{O}_i^d \quad (10.2)$$

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda_{\text{NP}}} C_1^{(5)} \mathcal{O}_1^{(5)} + \frac{1}{\Lambda_{\text{NP}}^2} \sum_i C_i^{(6)} \mathcal{O}_i^{(6)} + h.c. + \dots \quad (10.3)$$

The higher-dimensional terms are suppressed by $\Lambda_{\text{NP}}^{d-4}$, where d is the dimension. At dimension 5 there is only one operator after gauge symmetry constraints have been applied. It violates lepton number and has no relation to top physics and so is not considered further. At dimension 6 however, there are 16 operators which are relevant for top quark production and decay [181]. These operators can alter the normalisation and/or shape of top quark kinematic distributions. One particular operator of interest is \mathcal{O}_{tG} which modifies the $t\bar{t}g$ vertex and introduces a new $t\bar{t}gg$ vertex, as shown in Fig. 88, and results in a modification to the $t\bar{t}$ production cross section. The absolute differential cross section measurements presented in this thesis have sensitivity to this modified $t\bar{t}$ production cross section.

The EFT calculations are shown with respect to the H_T event variable in the rest of this section, however similar distributions can be found with respect to the other event variables in App. L. A LO model for the SM and top quark related EFT operators, including the \mathcal{O}_{tG} operator, is detailed in [182]. Constraints on the parameter $C_{tG}/\Lambda_{\text{NP}}^2$ can be calculated following prescriptions given in [165, 182]. Events based on the EFT model are generated using

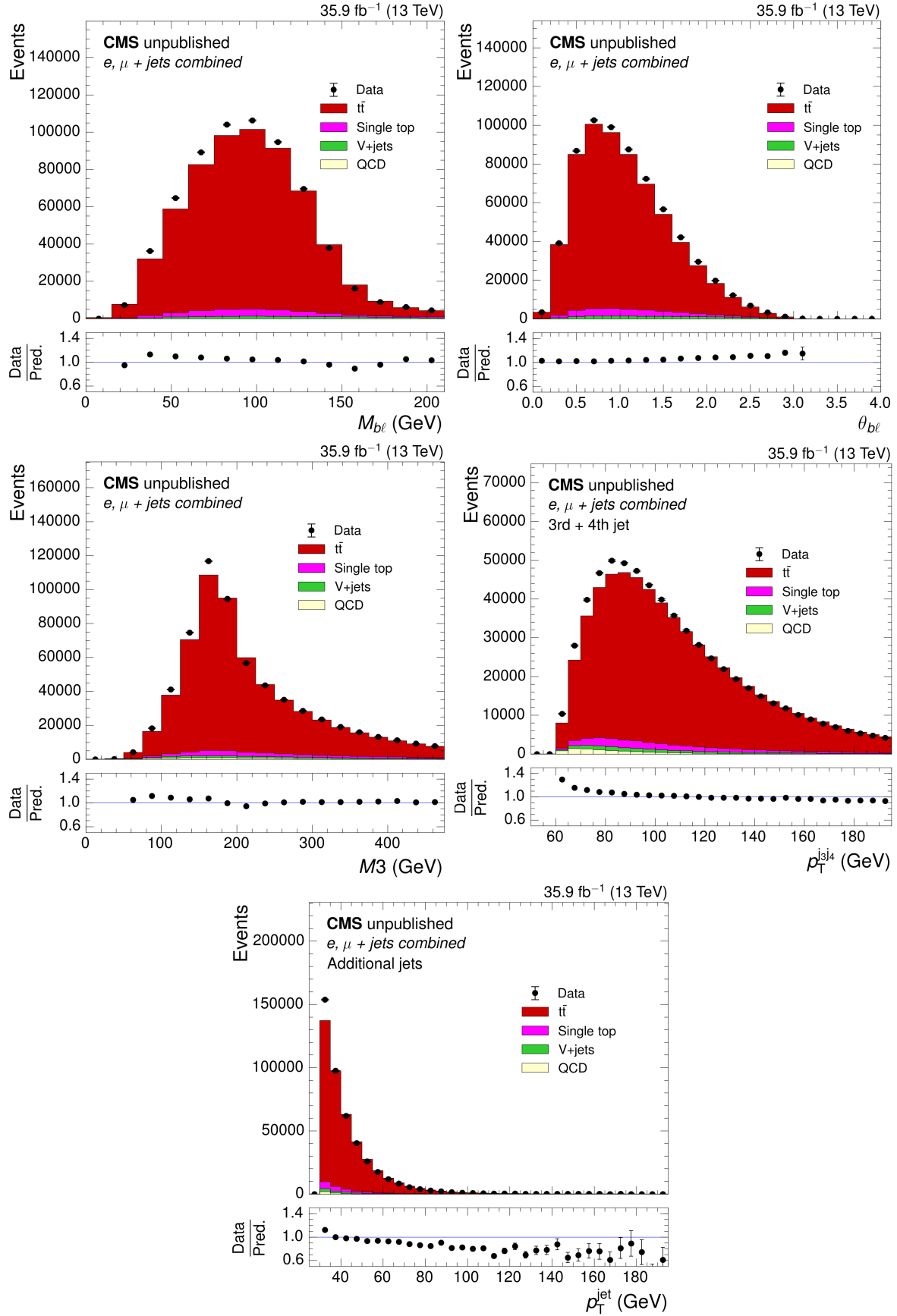


Figure 87: The distributions of $M_{b\ell}$, $\theta_{b\ell}$, $M3$, p_T^{j3j4} and additional p_T^{jet} after full event selection. The $t\bar{t}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars.

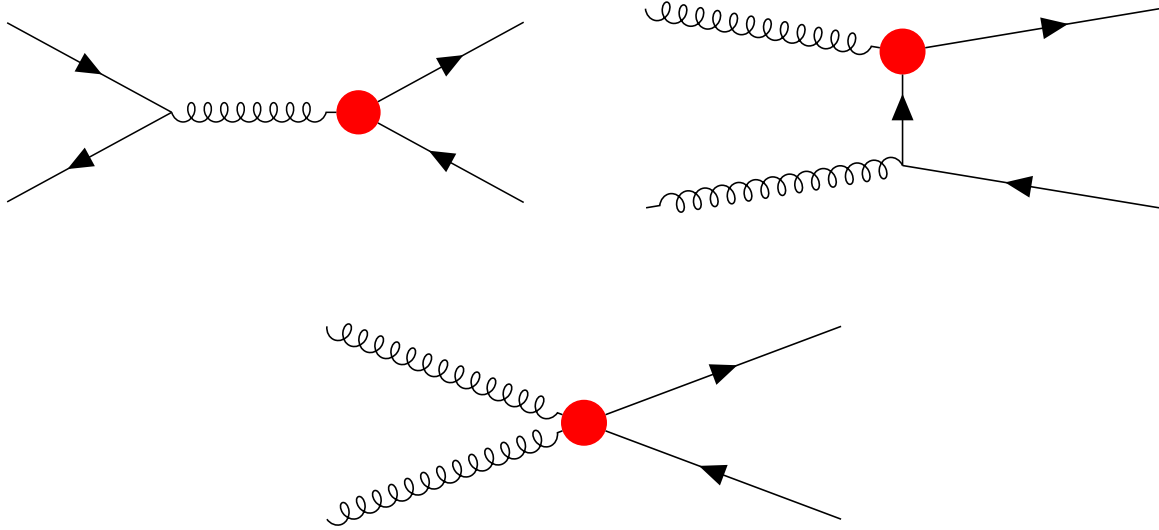


Figure 88: The upper two panels show the Feynman diagrams for the modified $t\bar{t}g$ vertices and the lower panel the additional $t\bar{t}gg$ vertex from the inclusion of the \mathcal{O}_{tG} operator. The red dot indicates the effective vertex.

MG5_aMC@NLO (v2.6.2) interfaced with PYTHIA (v8.238). It is not feasible, however, to generate many different simulated samples depending on different strengths of C_{tG}/Λ_{NP}^2 . Instead, the interference contributions can be scaled to the desired value of C_{tG}/Λ_{NP}^2 using the following method. The $t\bar{t}$ production cross section can be written as a perturbative expansion in terms of C_{tG}/Λ_{NP}^2

$$\sigma_{t\bar{t}} = \sigma_{t\bar{t}}^{\text{SM}} + \frac{C_{tG}}{\Lambda_{NP}^2} \cdot \beta_1 + \left(\frac{C_{tG}}{\Lambda_{NP}^2} \right)^2 \cdot \beta_2, \quad (10.4)$$

where the β terms are \mathcal{O}_{tG} contributions to the cross section from linear and quadratic perturbations of C_{tG}/Λ_{NP}^2 . By taking arbitrary values for this perturbation at $\pm X$ this reduces to

$$\beta_1 = \frac{\sigma_{t\bar{t}}(+X) - \sigma_{t\bar{t}}(-X)}{2X}, \quad (10.5)$$

which is the linear interference contribution for the Feynman diagrams containing one SM vertex and one \mathcal{O}_{tG} vertex. The interference term is then scaled to the desired C_{tG}/Λ_{NP}^2 strength and added to the SM prediction ($X = 0$) for the final EFT prediction. The quadratic contributions to the total $t\bar{t}$ production cross section are negligible for $C_{tG}/\Lambda_{NP}^2 < 1 \text{ TeV}^{-2}$ [183].

Three EFT simulations are created using coupling strengths of $C_{tG}/\Lambda_{NP}^2 = +1, 0, -1 \text{ TeV}^{-2}$. An independent sample of LO SM $t\bar{t}$ production is also simulated using the same generator in order to provide a scaling from LO to NNLO which is applied to all three EFT simulations. The scaling is calculated to be 1.69 and applied to all distributions. The three EFT simulations are shown in the upper panel of Fig. 89 and compared, using the RIVET plugin, to the absolute $t\bar{t}$ production cross section with respect to the H_T event variable, as measured in the measurements presented in this thesis.

The best-fit C_{tG}/Λ_{NP}^2 strength can then be extracted from a fit of χ^2 goodness-of-fit tests between the absolute cross sections and predictions over a given a range of C_{tG}/Λ_{NP}^2 . The goodness-of-fit tests follow an identical method to those described in Sec. 9.2 and use the covariance matrix of the absolute cross section measurements. The lower panel of Fig. 89 shows nominal fit of the $\Delta\chi^2$ value (the difference between the χ^2 value and the best fit χ^2 value) for different C_{tG}/Λ_{NP}^2 strengths when considering the absolute $t\bar{t}$ production cross sections with respect to the H_T event variable. The C_{tG}/Λ_{NP}^2 strength at the best fit χ^2 value is shown surrounded by the 68% *confidence interval* (CI) and the 95% confidence interval. The best fit

value for $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ is found to be -0.78 TeV^{-2} with a 95% CI $-1.86 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.29 \text{ TeV}^{-2}$. A summary table of the resulting CIs from all the other event variables is shown in App. L.

This basic study provides large confidence intervals already consistent with current measurements, for example $-0.06 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.41 \text{ TeV}^{-2}$ in [165] which is calculated using a NLO EFT model. Many aspects of this study can, and need, to be improved, such as the implementation of a NLO model using an alternative generator to MG5_aMC@NLO-LO+PYTHIA and the combination of the measurements from the different event variables, which involves calculating the correlations between the event variables. In addition, the scaling of the EFT simulations from LO to NNLO using SM simulations may not be valid in an EFT context.

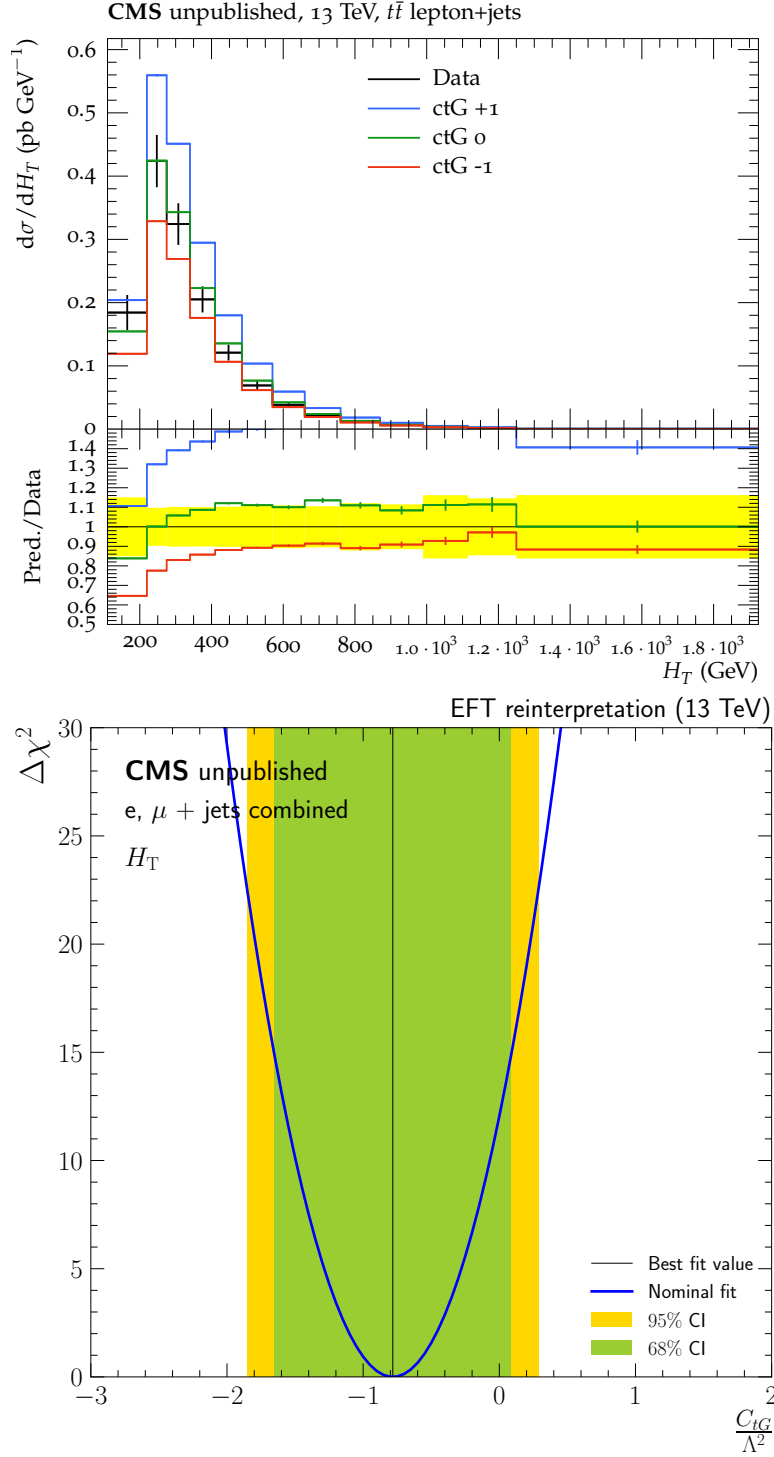


Figure 89: The upper panel shows the absolute differential cross sections with respect to the H_T event variable, predicted by setting the $c_{tG}/\Lambda_{\text{NP}}^2$ parameter to -1 (red), 0 (green) and +1 (blue) TeV^{-2} respectively, compared to that measured in this thesis. The lower panel shows the $\Delta\chi^2$ value for EFT models of differing $c_{tG}/\Lambda_{\text{NP}}^2$ strengths and the absolute $t\bar{t}$ cross section with respect to the H_T event variable. The nominal fit (blue) is shown together with its best fit value (black) and 68% (green) and 95% (gold) confidence intervals.

Summary

The field of particle physics sets out to answer two of the greatest, most fundamental questions of our age. What are the constituents of the Universe, and how do they interact? At its heart is the SM, unparalleled in its success at predicting the existence of and confirming all of the observed particles and their interactions. It is not without problems though, with its inability to explain gravity or to describe the baryonic content of the Universe. These, and other issues, motivate the work of particle physicists determined to create a complete picture of our Universe.

The top quark is the most massive, elementary particle of the SM, and the study of which is expected by particle physicists to be the most sensitive to any BSM physics. This new physics could directly manifest itself in the production, decay and couplings of the $t\bar{t}$ system. It will decay without forming a hadron, allowing physicists to study its bare quark properties through its decay products. The top quark also appears in, and can be a major background to, many rare SM and BSM processes. For these reasons, and more, precision measurements of the properties of the top quark, especially inclusive and differential cross sections, are vital.

This thesis presents measurements of differential $t\bar{t}$ production cross sections as a function of several kinematic event variables using events with a single electron or muon. Kinematic event variables do not require the reconstruction of the complete $t\bar{t}$ system. The event variables considered are the jet multiplicity, the scalar sum of the jet p_T and the scalar sum of the p_T of all particles, the magnitudes of the p_T imbalance and the p_T of the leptonically decaying W boson and finally, the magnitudes of the p_T and η of the lepton. This set of event variables is extremely useful. Some are sensitive to not just the hard interaction, but also the soft processes underneath, making them very suitable in the validation of event generation and modelling. Many of the event variables are also used as the search variable of choice when looking for BSM physics, so precise measurements of the differential cross section for, often, the largest background, with respect to these search variables is very useful. The event variables can also be used as proxies for the top quark, without having to reconstruct it, providing additional information from a subtly different (metaphorical) angle.

The measurements are performed using collision data collected by the CMS detector at the LHC in 2016 and the total amount of data used in this thesis corresponds to 35.9 fb^{-1} . Normalised and absolute differential cross sections are calculated to particle level in a visible phase space and compared, using goodness-of-fit tests, to four state-of-the-art $t\bar{t}$ production models:

- POWHEG+PYTHIA
- POWHEG+HERWIG++

- MG5_aMC@NLO-NLO+PYTHIA
- MG5_aMC@NLO-LO+PYTHIA

The NLO models are found to be consistent with the data, however the MG5_aMC@NLO-LO+PYTHIA generator is not. Tension was observed in the measurement with respect to the $|\eta^\ell|$ event variable for all models, which originates from the μ +jets decay channel. It is thought to come from shape effects, possibly introduced with the application of scale factors, but more investigation is necessary.

These cross section measurements are a part of a suite of 13 TeV complementary measurements, across different final states, phase spaces and experiments [3, 87–90, 98–102]. In addition, these measurements can be used for constraints on the PDF of the proton, as an accurate background estimate for BSM or rare SM searches and as an input to phenomenology studies. They have been made available to the wider community by implementing them in RIVET and are already being used in the tuning and validation of the next generation of $t\bar{t}$ models.

Final remark from the author

Due to the nature of the top quark, the field of top quark physics is by necessity a wide one. Among the vast array of topics, differential cross section measurements of the top quark pair production will always be relevant and important, especially so, as the next generation of particle colliders are considered. At higher collision energies and luminosities, hints of new physics in the top sector seem just around the corner. In the future electron-positron colliders, the top quark properties can be measured to an unprecedented degree of precision, and in doing so, test the compatibility of the SM as rigorously as possible. The field of top quark physics is integral to the field of particle physics and it has been one that I’ve loved to work in and I believe its future is very bright indeed.

Samenvatting

De deeltjesfysica tracht twee van de grootste, meest fundamentele vragen van dit tijdperk te beantwoorden. Wat zijn de bouwstenen van het Universum en hoe interageren deze bouwstenen? Centraal in deze vragen staat het standaard model, ongeëvenaard succesvol in de voorspelling en beschrijving van alle waargenomen deeltjes en hun interacties. Het standaard model is niet zonder gebreken, het is bijvoorbeeld niet in staat om zwaartekracht of niet-barionische materie te beschrijven. Deze en andere gebreken zijn het gevolg van deeltjes fysici die gedreven zijn om een zo compleet mogelijk beeld van het Universum te schetsen.

Er wordt verwacht dat het top quark, het zwaarste deeltje in het standaard model, het meest gevoelig is aan nieuwe fysica voorbij het standaard model. Deze nieuwe fysica toont zichzelf rechtstreeks in de productie werkzame doorsnede van het top quark-antiquark paar ($t\bar{t}$). Om deze en andere redenen is een precieze kennis van de inclusieve en partiële werkzame doorsneden van het standaard model top quark uitermate belangrijk.

Deze thesis presenteert een meeting van de differentiële $t\bar{t}$ productie werkzame doorsnede in functie van verschillende kinematische evenement variabelen waarbij evenementen met een enkel elektron of muon gebruikt worden. Kinematische evenement variabelen zijn variabelen waarvoor de volledige reconstructie van het compleet $t\bar{t}$ systeem niet vereist zijn. De evenement variabelen onder beschouwing zijn jet veelvuldigheid, de scalaire som van de transversale jet impulsen, de scalaire som van de transversale impulsen van alle deeltjes, de grootte van het transversaal impuls onevenwicht, de transversale impuls van het leptonisch vervallend W en uiteindelijk ook de grootte van de transversale impuls en de pseudorapiditeit van het lepton.

De metingen werden uitgevoerd met botsingsdata verzameld met de CMS detector aan het LHC, de totale hoeveelheid data gebruikt in deze thesis komt overeen met 35.9 fb^{-1} . Evenementen waarbij zowel een enkel geïsoleerd elektron of muon voorkomt als minstens vier jets waarvan er minstens twee geïdentificeerd zijn als afkomstig van een b quark worden beschouwd. Bijkomstige selecties op de kwaliteit van de leptonen en de jets werden toegepast. De $t\bar{t}$ opbrengst wordt geschat door zowel V +jets en enkelvoudige top achtergrond simulaties als QCD bijdragen bepaald met data van de data af te trekken. De metingen worden voorgesteld op deeltjes niveau waarbij de kinematische distributies opgebouwd worden uit stabiele deeltje (deeltjes met een gemiddelde levensduur van meer dan 30 ps) in de detector. Hierbij wordt een faseruimte gebruikt die gelijkaardig is aan de faseruimte beschikbaar voor de detector. Het aantal evenementen die gereconstrueerd worden door de detector maar die zich niet in de zichtbare faseruimte bevinden worden afgetrokken van de $t\bar{t}$ opbrengst. Hierna wordt de opbrengst ontvouwd om de effecten van de detector acceptantie en efficiëntie en bewegingen tussen bins door de detector resolutie in rekening te brengen.

Zowel de genormaliseerde als de absolute werkzame doorsnede worden berekend ten opzichte van de ontvouwde opbrengst. Ze worden ook vergeleken met de beste $t\bar{t}$ modellen: POWHEG+PYTHIA, POWHEG+HERWIG++, MG5_aMC@NLO-LO+PYTHIA en MG5_aMC@NLO-NLO+PYTHIA. Fit kwaliteit tests tussen de gemeten en de gegenereerde werkzame doorsneden werden uitgevoerd. Er werd bevonden dat powhegpythia over het algemeen consistent is met de data waarbij verschillen binnen de theoretische onzekerheden van de modellen vallen. De POWHEG+HERWIG++ en MG5_aMC@NLO-NLO+PYTHIA modellen zijn consistent met de data voor het grootste deel van de kinematische evenement variabelen, MG5_aMC@NLO-LO+PYTHIA beschrijft geen enkele kinematische variabele correct.

Er wordt verwacht dat deze metingen gebruikt zullen worden om de $t\bar{t}$ generatoren te verbeteren. In dat opzicht werden de resultaten ook in RIVET beschikbaar gemaakt voor de bredere gemeenschap. Bovendien kunnen deze metingen gebruikt worden om de parton distributie functie van het proton te beperken, als nauwkeurige achtergrond schatting voor zoektochten naar fysica voorbij het standaard model of zeldzame standaard model processen en als informatie voor fenomenologische studies.

Appendices

APPENDIX A

Purity and stability in the muon decay channel

Additional bias tests on the unfolding procedure are performed by unfolding $t\bar{t}$ events generated from an alternative model and comparing to the true values. These are shown in Figs. 90 and 91 in the e +jets and μ +jets channels respectively. For the most part, they lie comfortably within the systematic uncertainties, shown as the grey band in the ratio plots.

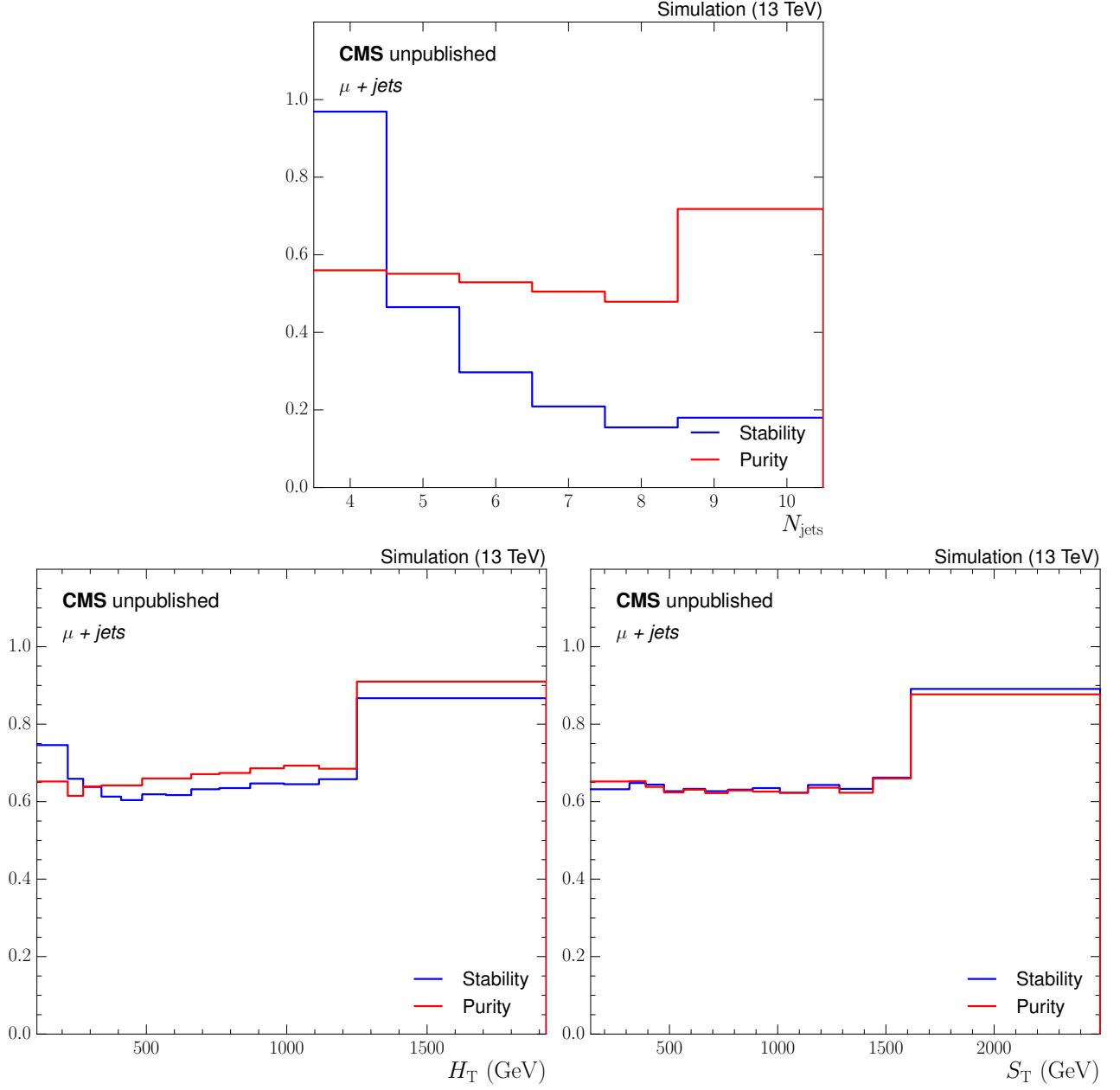


Figure 90: The purity and stability of the bins in the $\mu + jets$ channel, measured using the POWHEG+PYTHIA simulation sample, for the N_{jets} , H_T and S_T event variables.

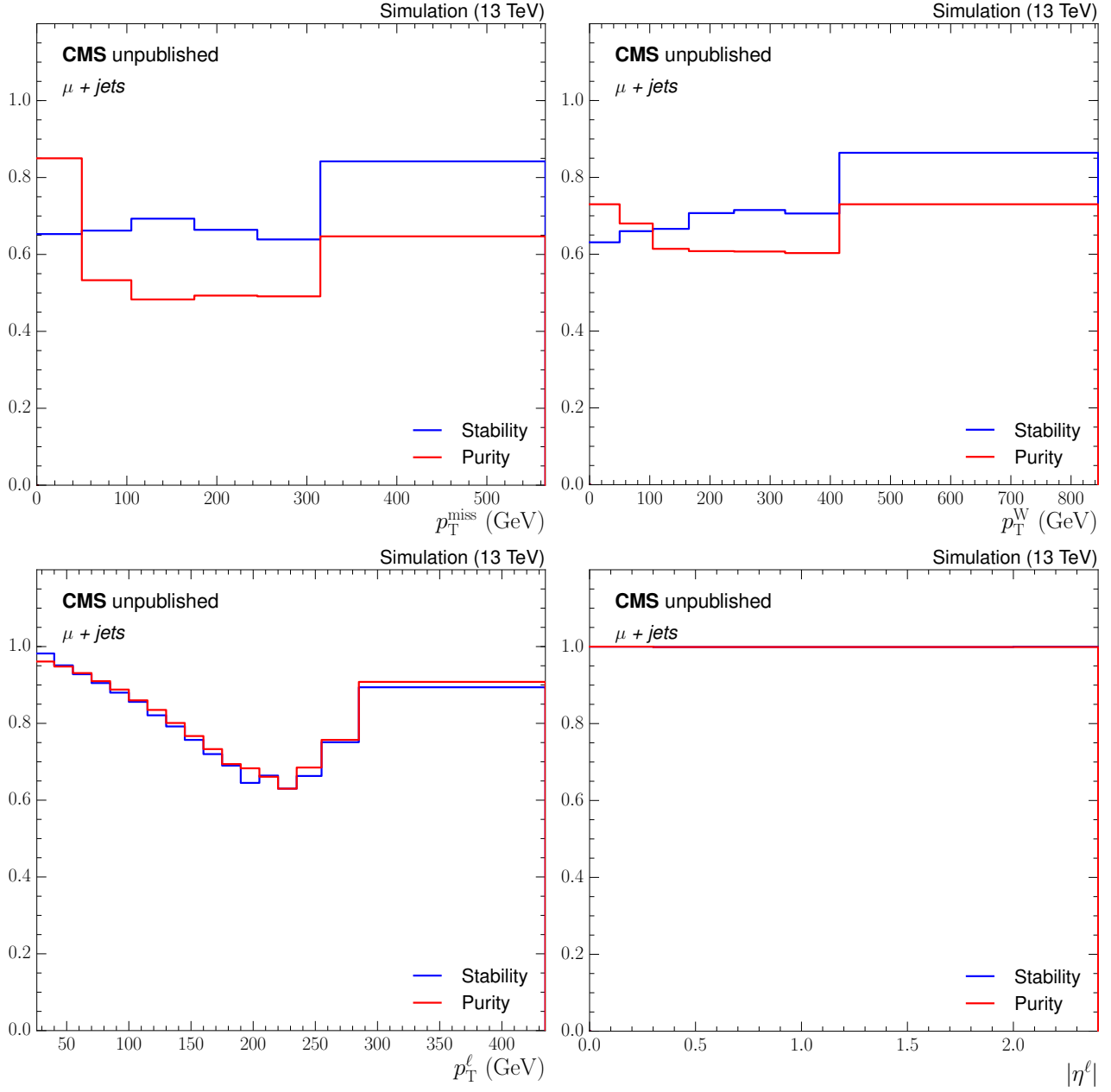


Figure 91: The purity and stability of the bins in the $\mu + jets$ channel, measured using the POWHEG+PYTHIA simulation sample, for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables.

APPENDIX B

Residual distributions

Figures 92, 93, 94, 95, 96, 97, 98, 99 and 100 show the comparisons of the bin widths and the resolutions (calculated from the residual distributions) in each bin for all event variables except N_{jets} . The residual distributions are calculated using simulated events from the POWHEG+PYTHIA model where the events from the e+jets and μ +jets channels have been combined.

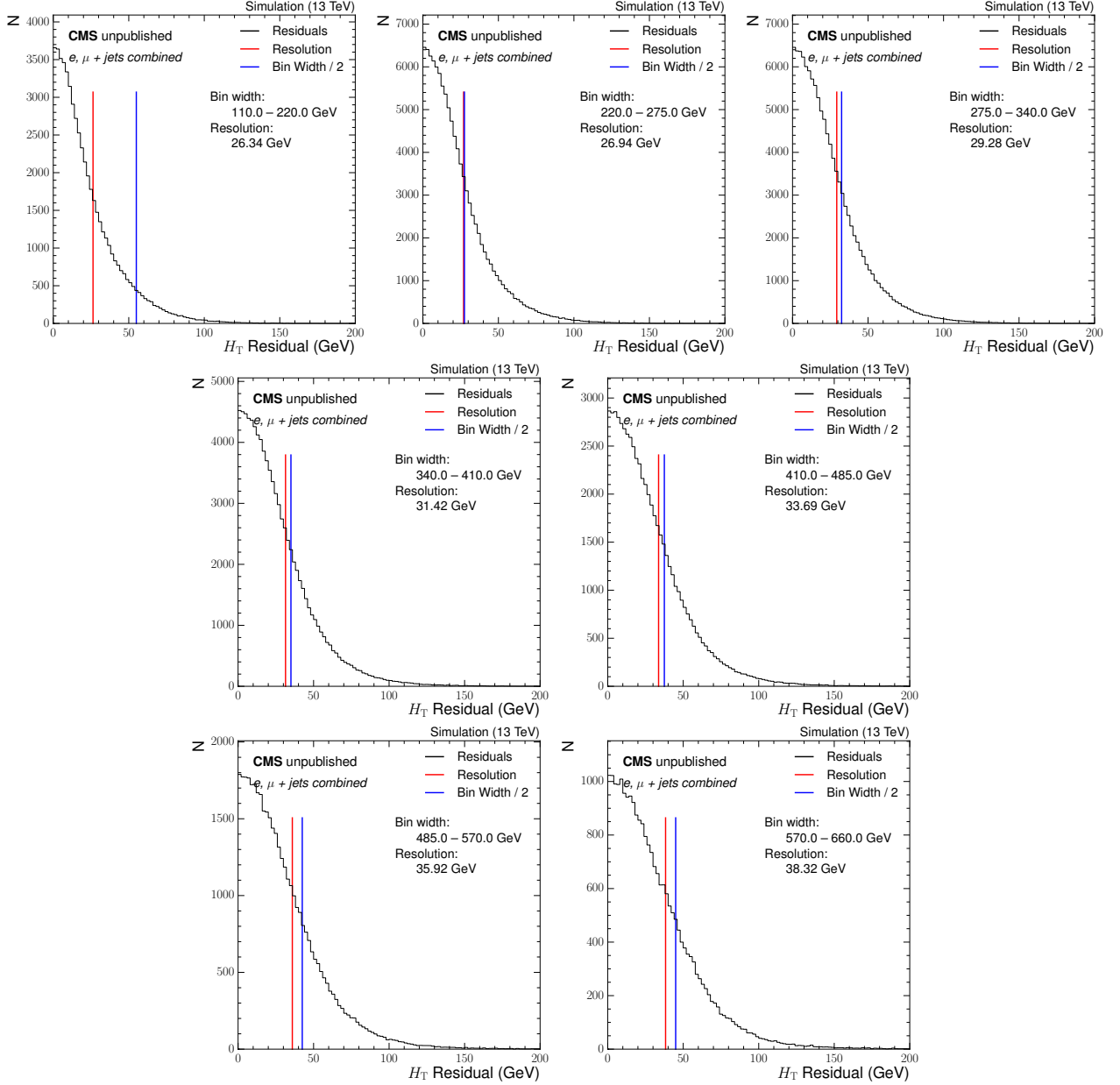


Figure 92: The set of residual distributions for the first 7 bins of the of the H_T event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

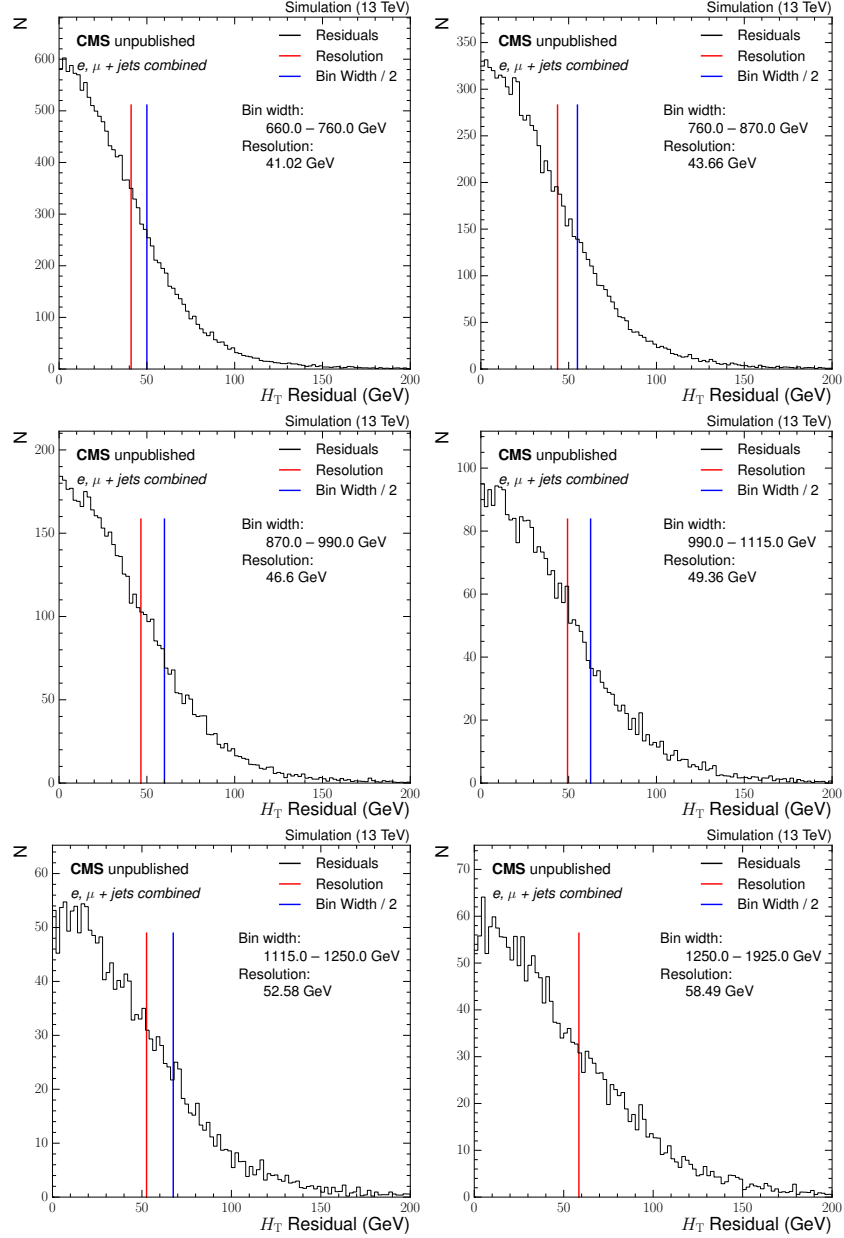


Figure 93: The set of residual distributions for the last 6 bins of the of the H_T event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

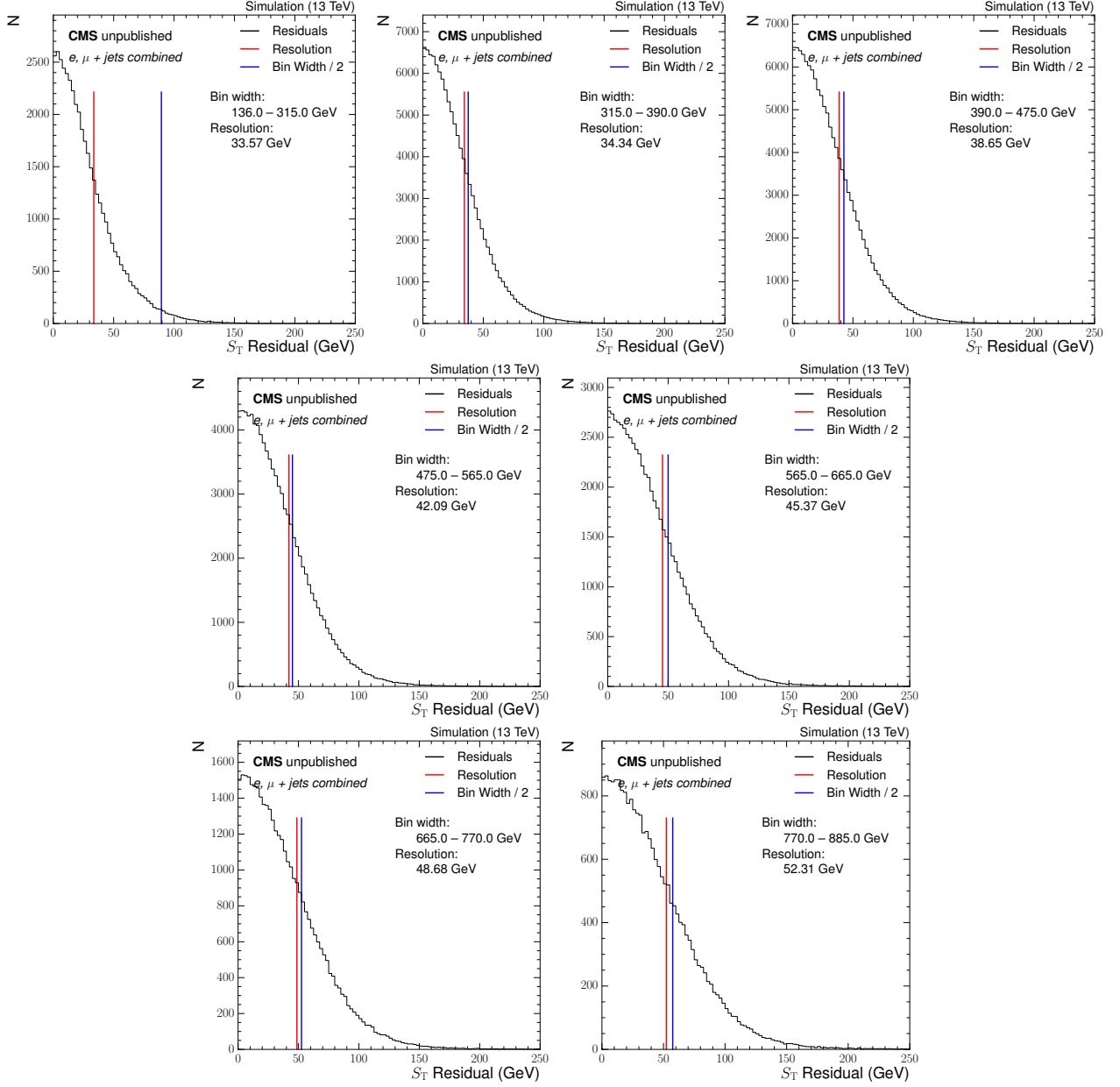


Figure 94: The set of residual distributions for the first 7 bins of the of the S_T event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

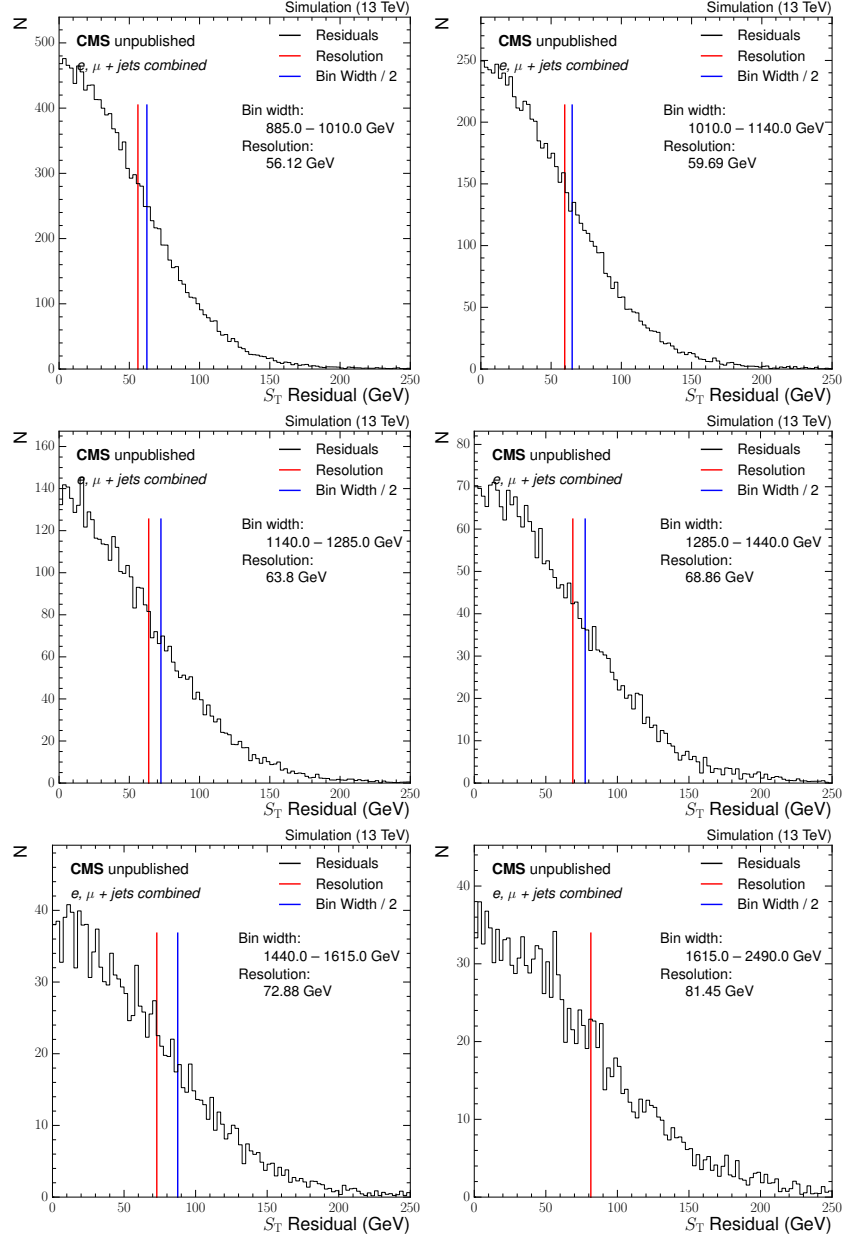


Figure 95: The set of residual distributions for the last 6 bins of the of the S_T event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

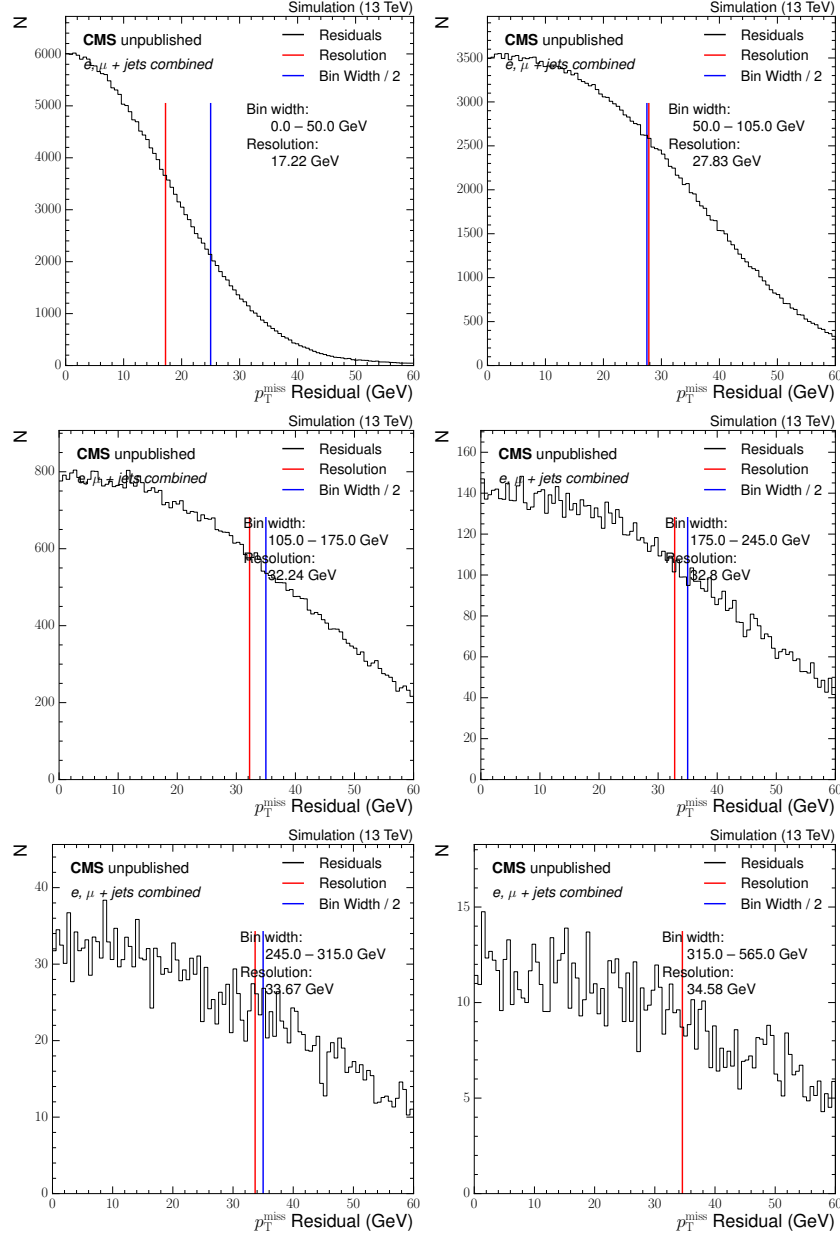


Figure 96: The set of residual distributions for the p_T^{miss} event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

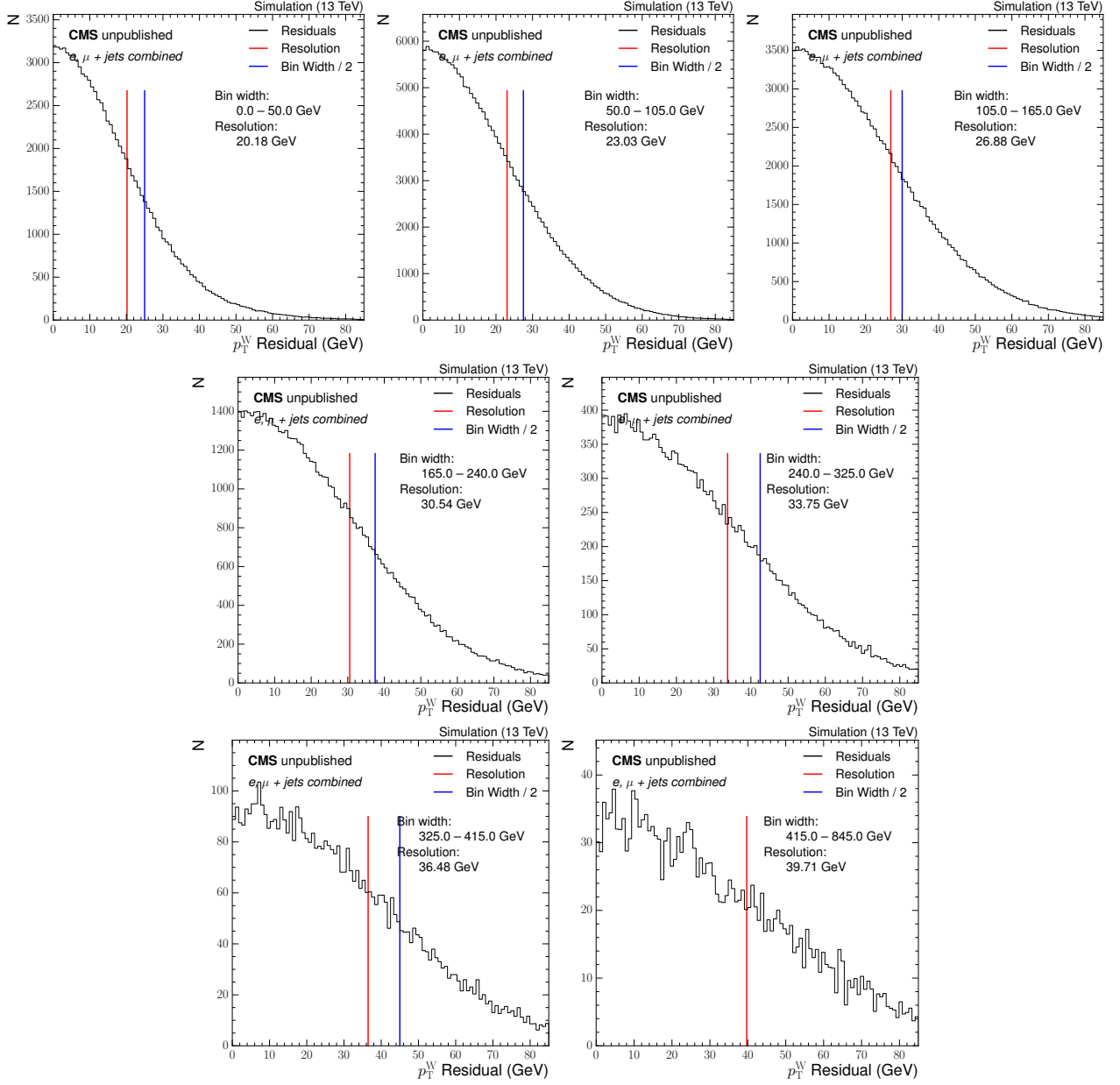


Figure 97: The set of residual distributions for the p_T^W event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

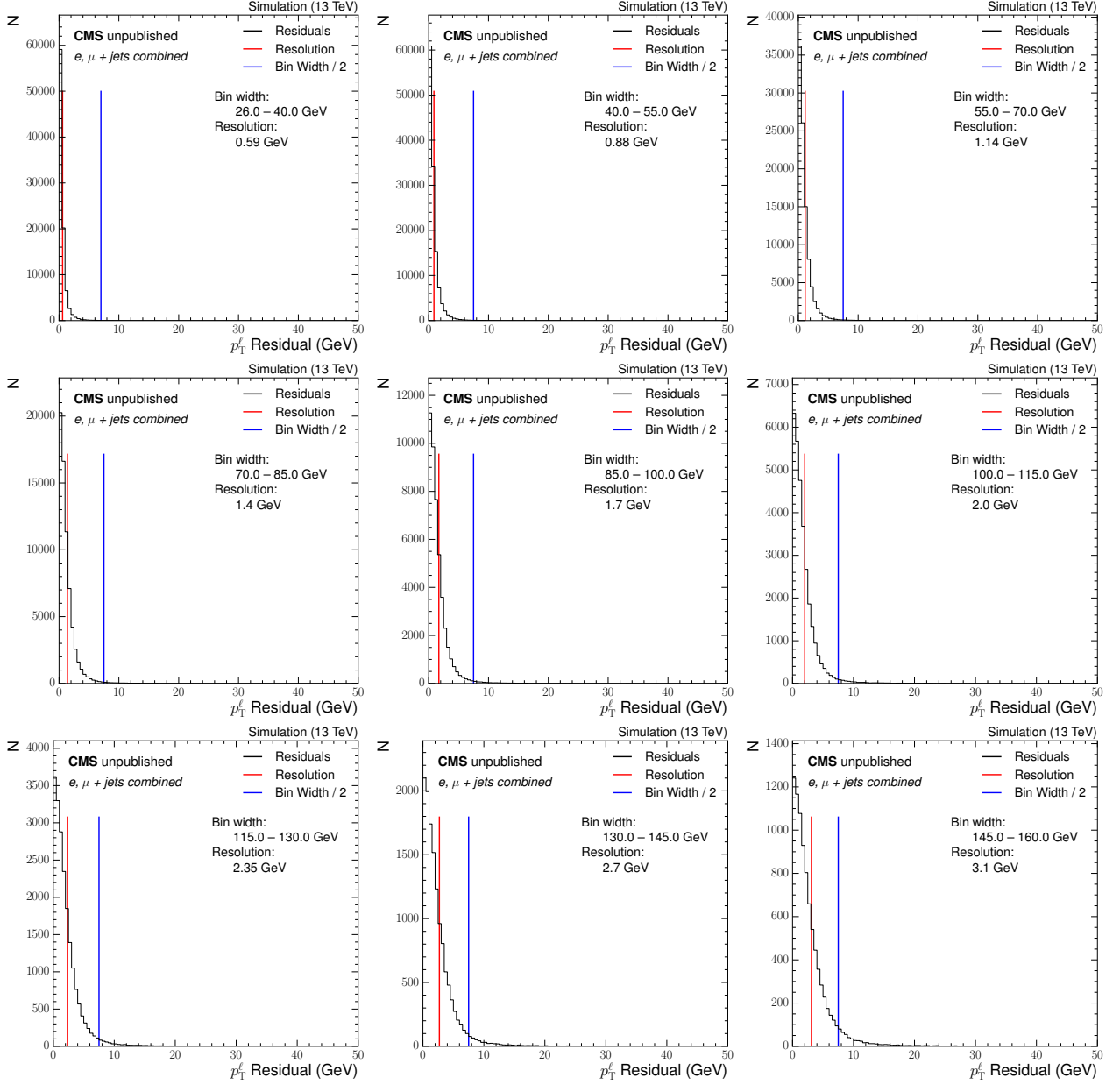


Figure 98: The set of residual distributions for the first 9 bins of the of the p_T^l event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

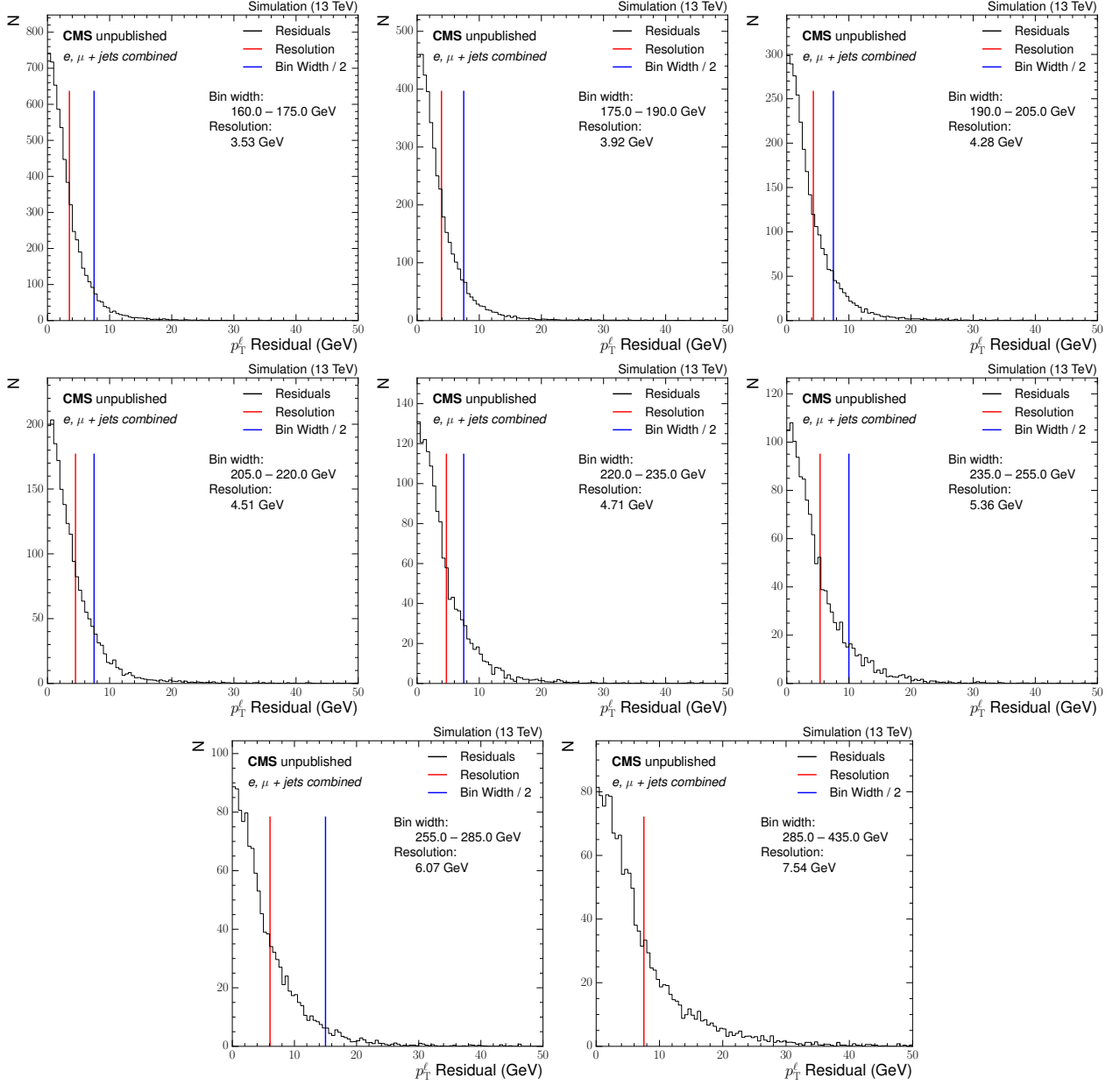


Figure 99: The set of residual distributions for the last 8 bins of the of the p_T^l event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

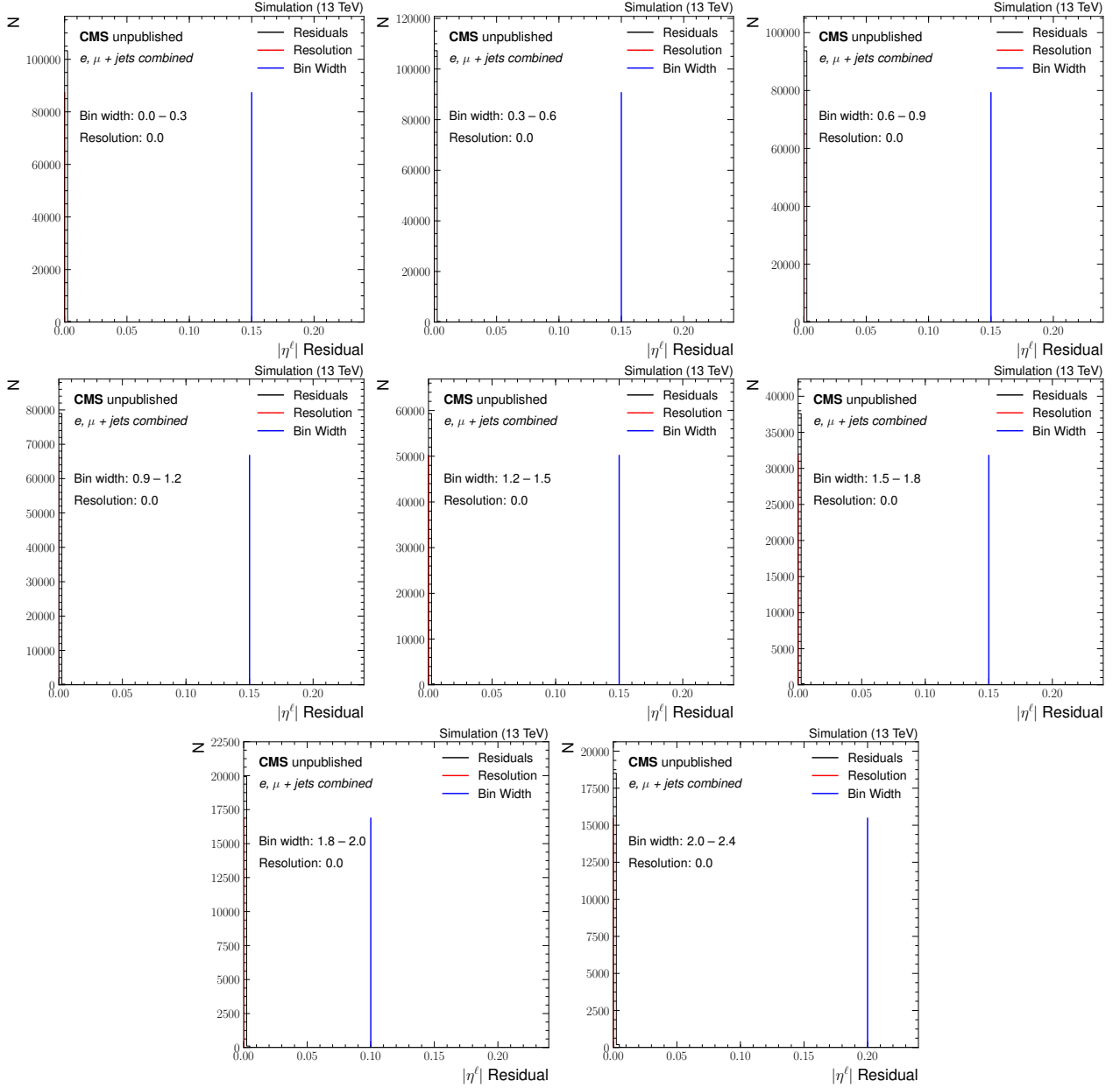


Figure 100: The set of residual distributions for the $|\eta^\ell|$ event variable calculated using the POWHEG+PYTHIA simulation. Overlaid is the half-bin width (red) and the resolution (blue)

APPENDIX C

Response matrices in the muon decay channel

The migration matrices, calculated using the POWHEG+PYTHIA model, for the μ +jets channel, calculated using the common binning scheme, are shown in Fig. 101 for the N_{jets} , H_{T} and S_{T} event variables and in Fig 102 for the $p_{\text{T}}^{\text{miss}}$, p_{T}^{W} , p_{T}^{ℓ} and $|\eta^{\ell}|$ event variables. The binning requirements result in close-to-diagonal migration matrices with an acceptable statistical uncertainty.

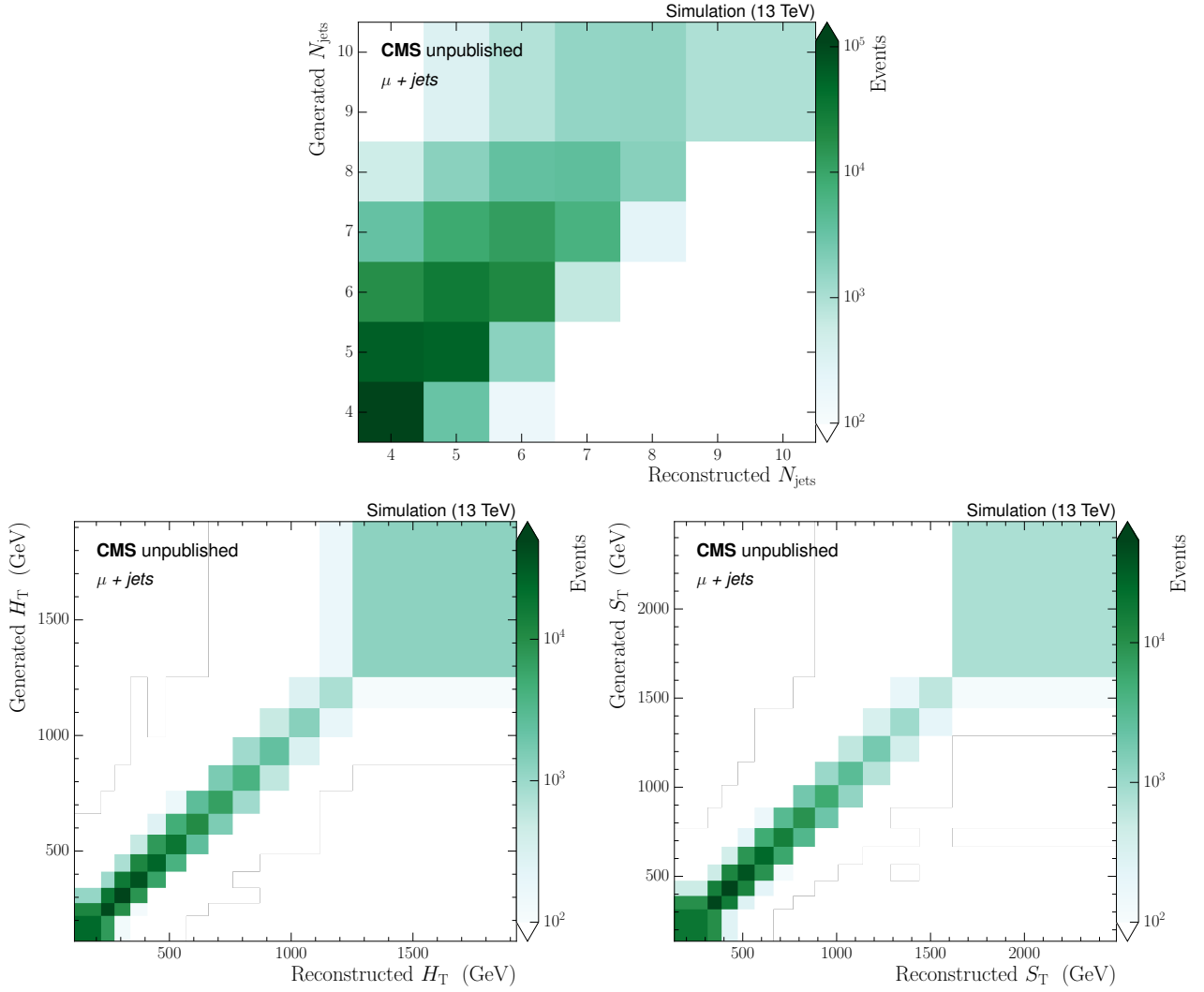


Figure 101: The set of migration matrices calculated using the POWHEG+PYTHIA simulation sample for the N_{jets} , H_T and S_T event variables in the $\mu+jets$ channel using the common binning scheme.

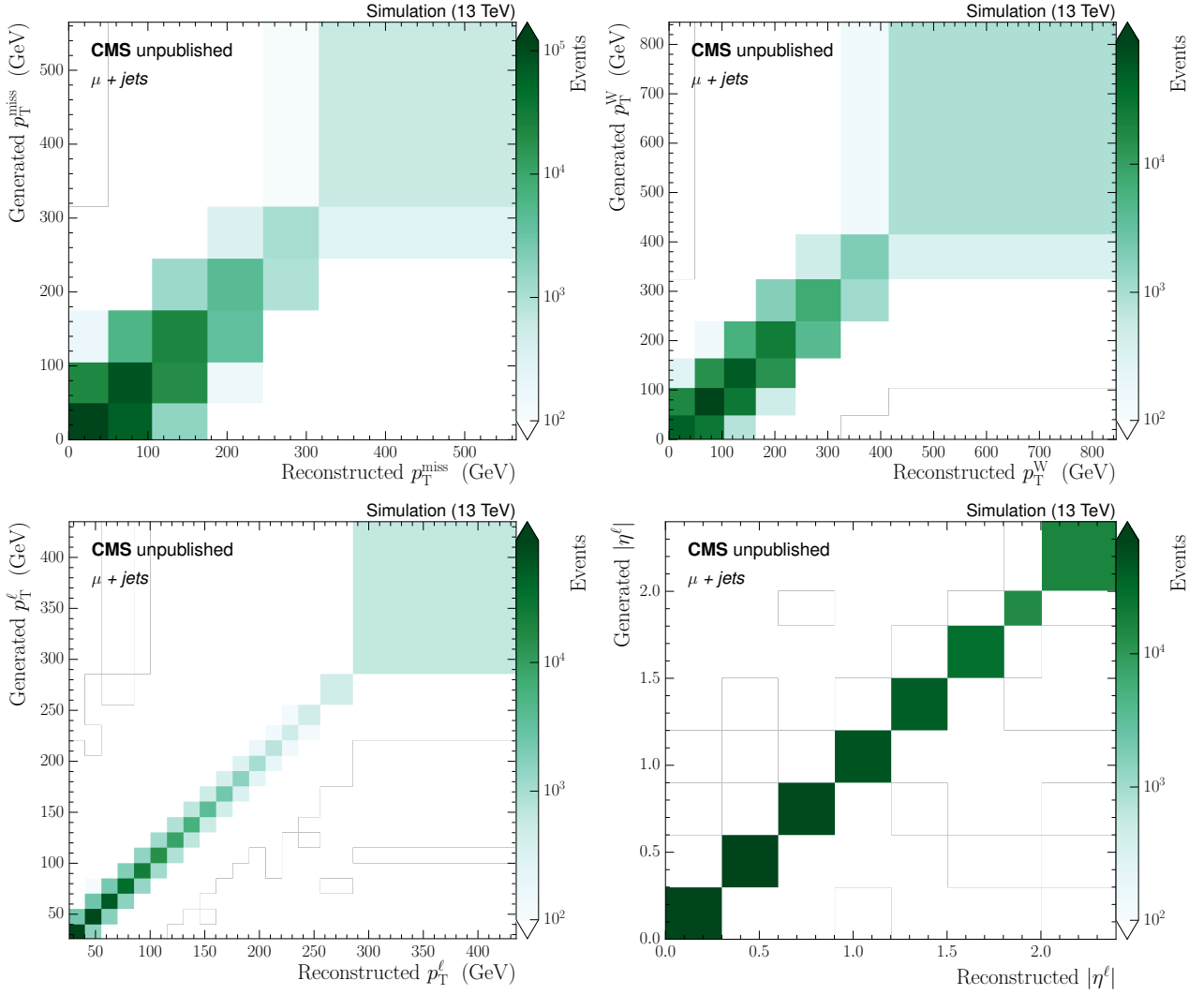


Figure 102: The set of migration matrices calculated using the POWHEG+PYTHIA simulation sample for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the $\mu + \text{jets}$ channel using the common binning scheme.

APPENDIX D

Regularisation strength in the muon decay channel

The regularisation strengths calculated by minimising the average global correlation coefficient of the bins, for the μ +jets channel, are shown in Fig. 103 for the N_{jets} , H_{T} and S_{T} event variables and in Fig 104 for the $p_{\text{T}}^{\text{miss}}$, p_{T}^{W} , p_{T}^{ℓ} and $|\eta^{\ell}|$ event variables.

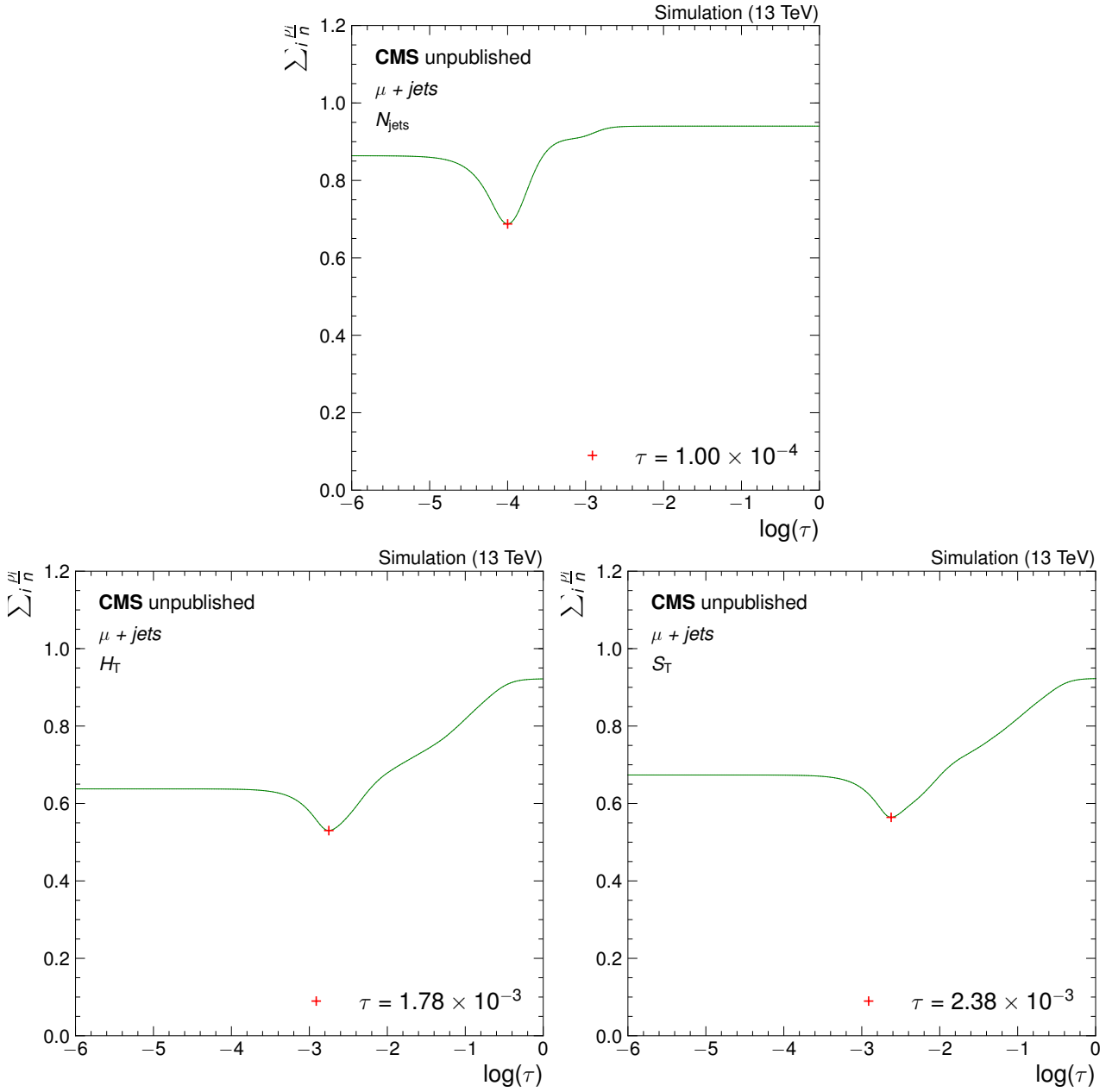


Figure 103: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the μ +jets channel, for the N_{jets} , H_T and S_T event variables. The optimal τ is shown as a red point at the minimum of the scan.

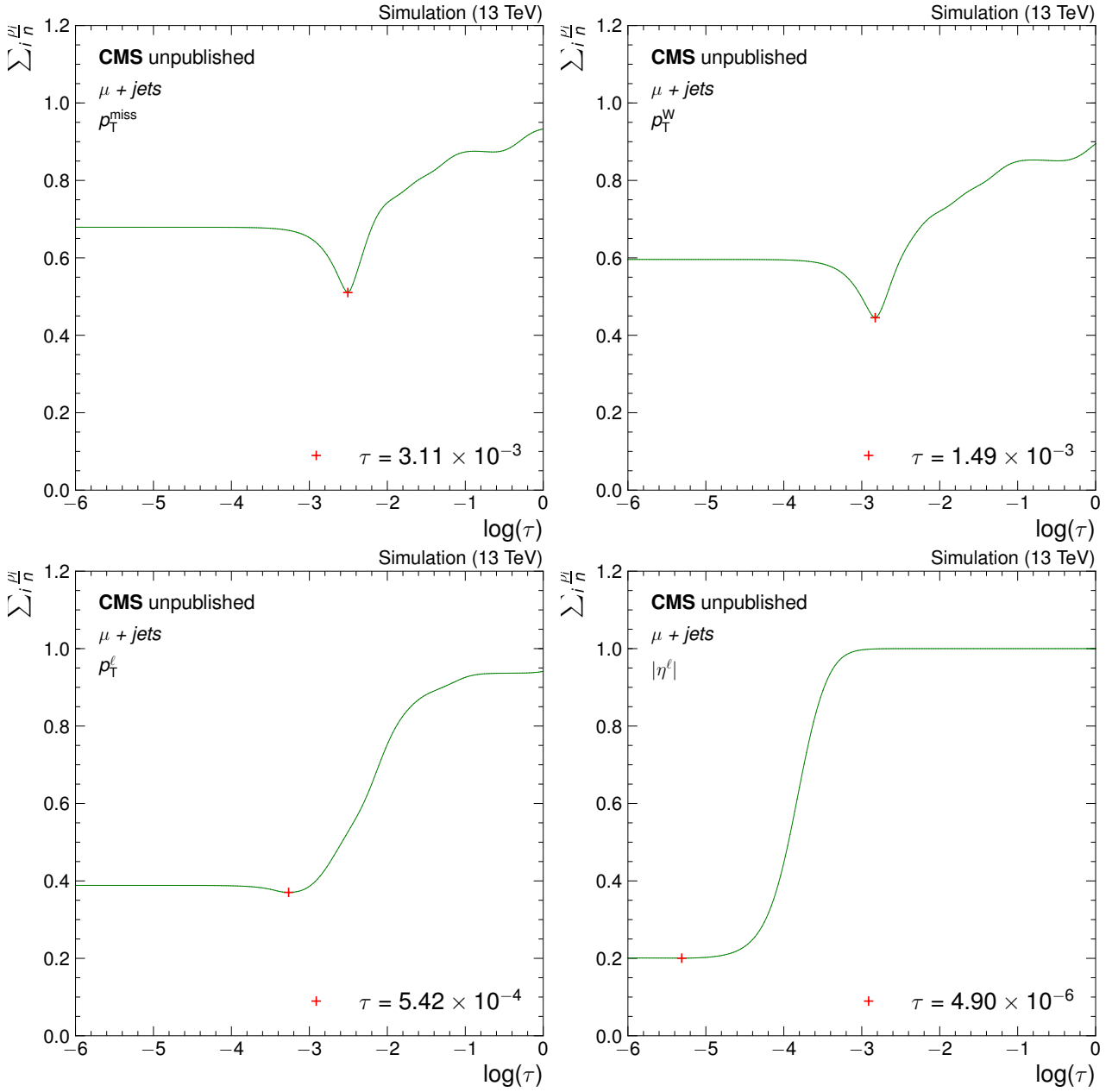


Figure 104: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the $\mu + jets$ channel, for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables. The optimal τ is shown as a red point at the minimum of the scan.

APPENDIX E

Pull distributions in the muon decay channel

The set of pull distributions for the μ +jets channel, are shown in Fig. 105 for the N_{jets} , H_{T} and S_{T} event variables and in Fig 106 for the $p_{\text{T}}^{\text{miss}}$, p_{T}^{W} , p_{T}^{ℓ} and $|\eta^{\ell}|$ event variables. It can clearly be seen that the mean and width of the pull distribution are close to zero and one, so the unfolding is treating the uncertainty and normalisation correctly.

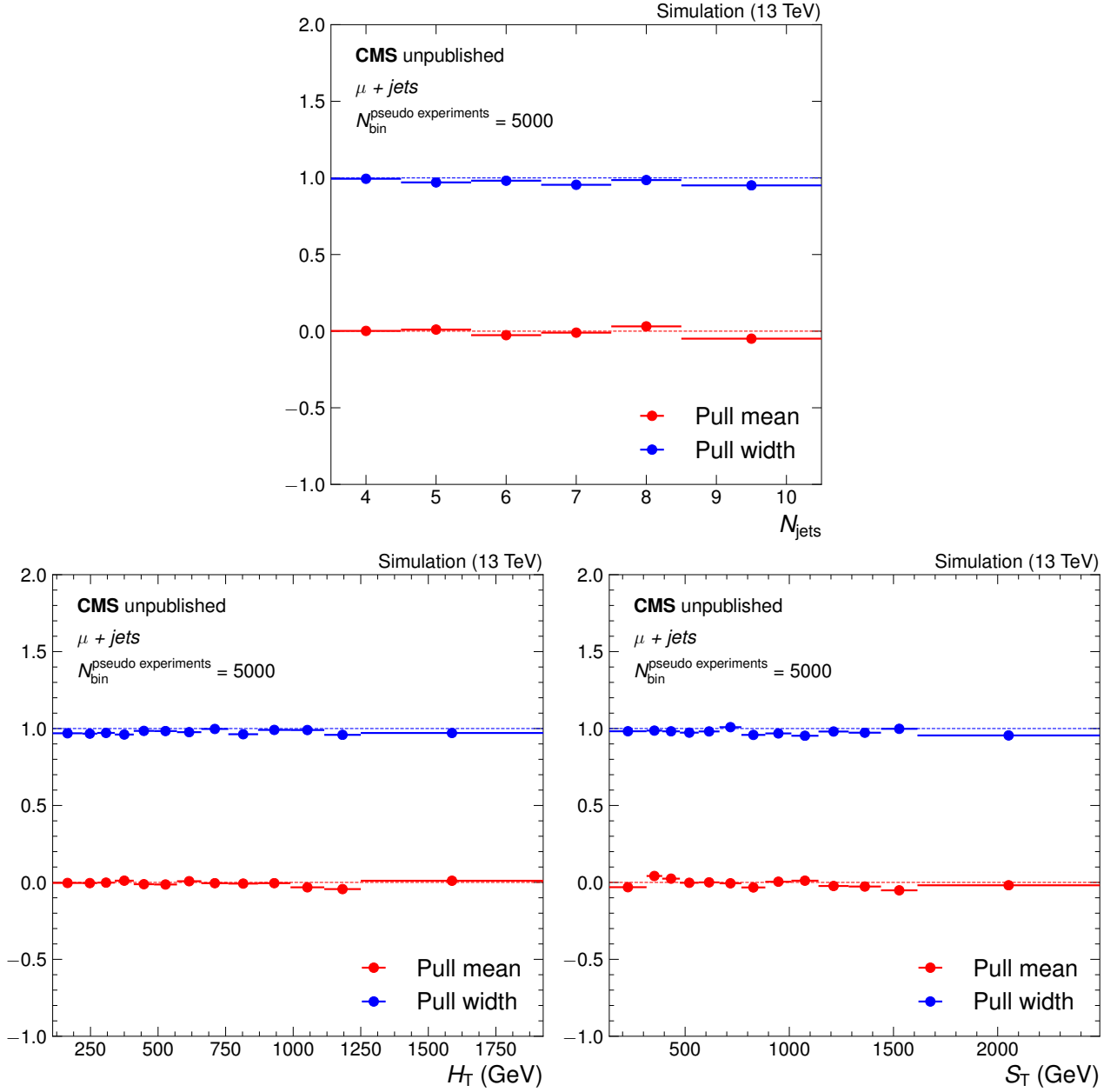


Figure 105: The pull mean and widths in relation to the bin numbers of the N_{jets} , H_T and S_T event variables in the $\mu+jets$ channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA simulation.

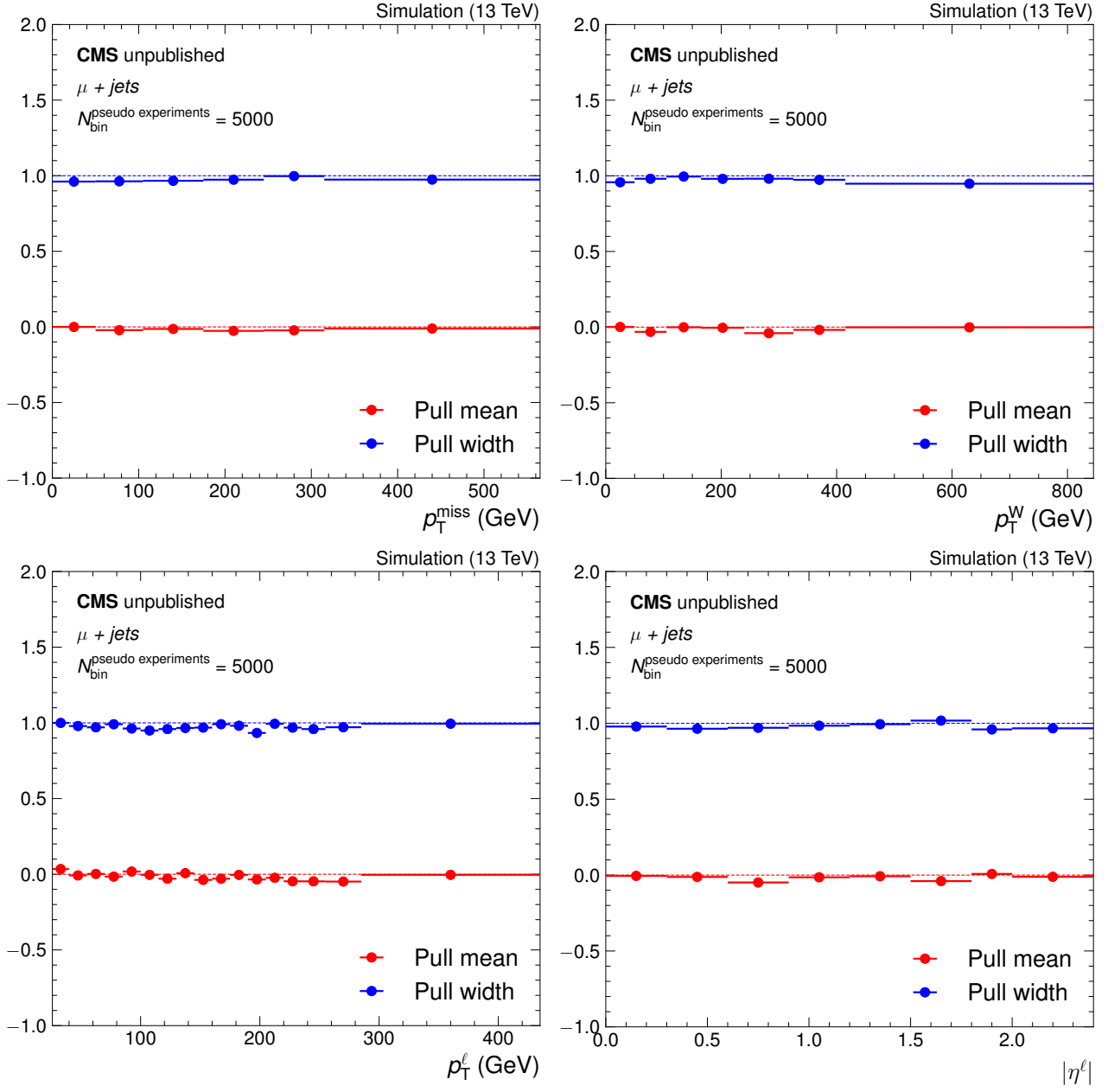


Figure 106: The pull mean and widths in relation to the bin numbers of the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the $\mu + jets$ channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA simulation.

APPENDIX F

Creating the envelopes to test for bias

The top quark p_T in the POWHEG+PYTHIA model is reweighted up and down to cover the differences between data and simulation in order to provide a basis for the bias tests. Figures 107, 108, 109 and 110 show the reweighted kinematic event distributions in the e+jets and μ +jets channels respectively.

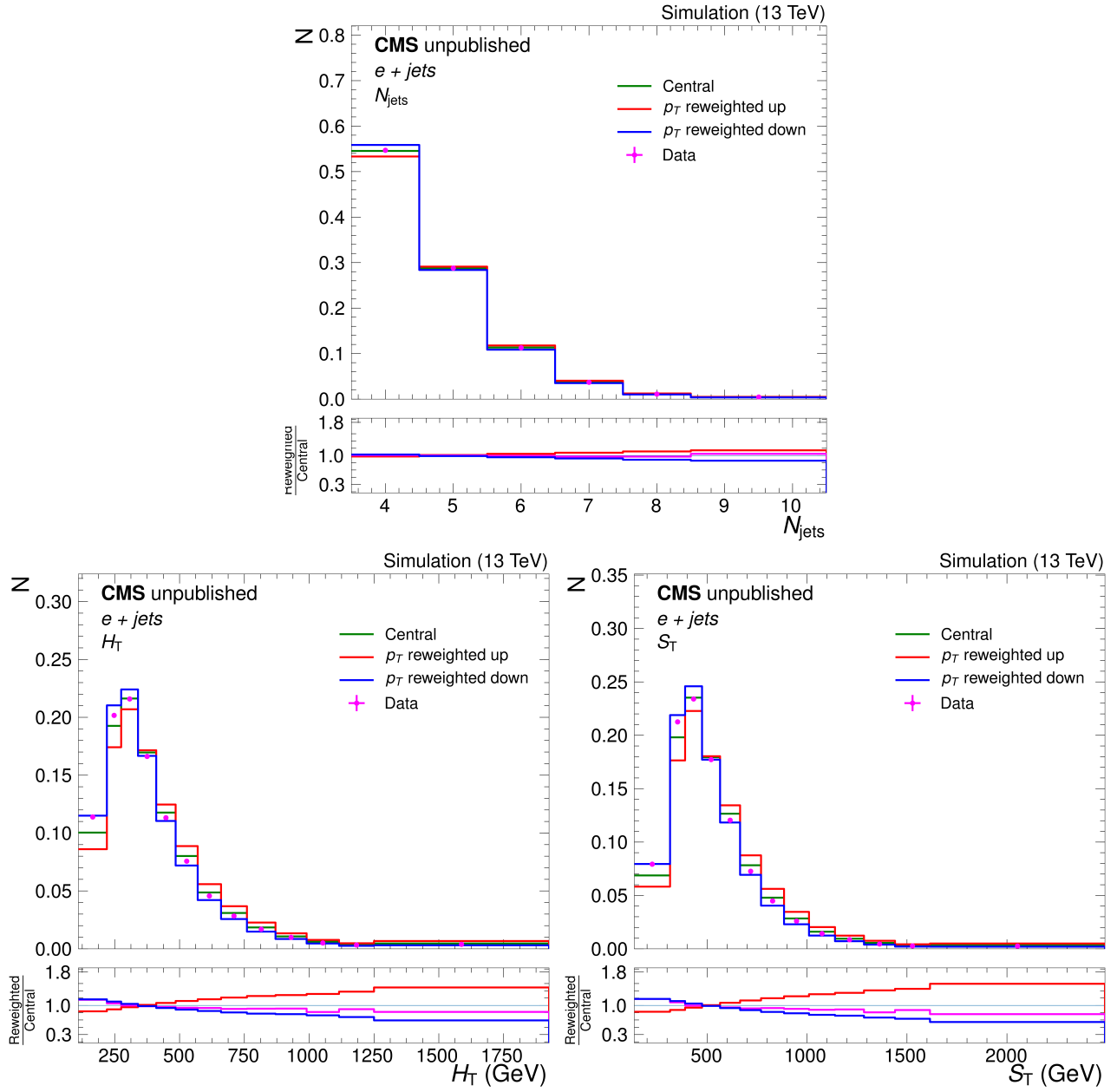


Figure 107: The kinematic event distributions given by the POWHEG+PYTHIA sample (green) with the top quark p_T reweighted up (red) and down (blue) to cover differences to data (magenta) in the $e+jets$ channel, for the N_{jets} , H_T and S_T event variables. The distributions are normalised to one.

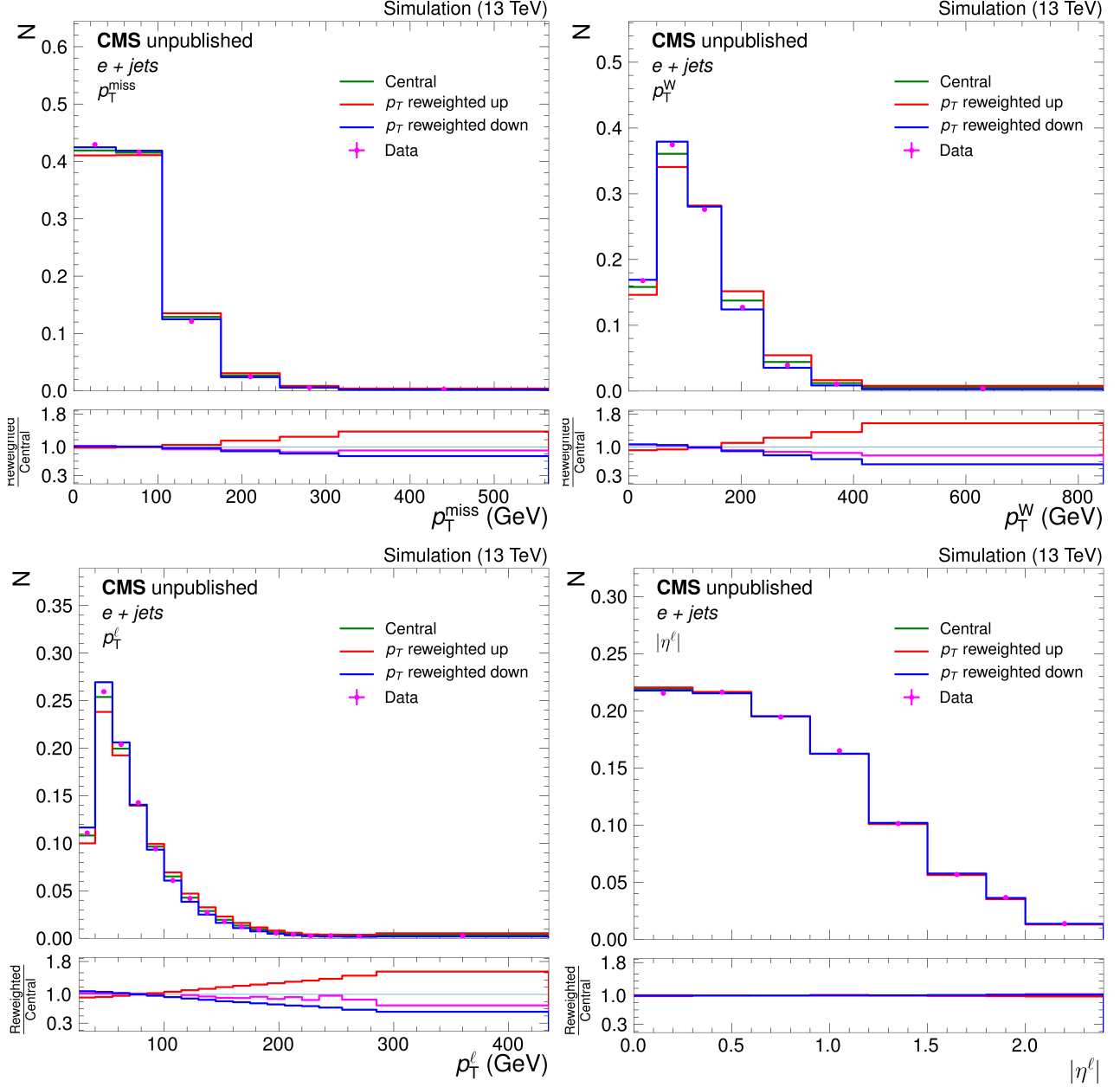


Figure 108: The kinematic event distributions given by the POWHEG+PYTHIA sample (green) with the top quark p_T reweighted up (red) and down (blue) to cover differences to data (magenta) in the $e + \text{jets}$ channel for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables. The distributions are normalised to one.

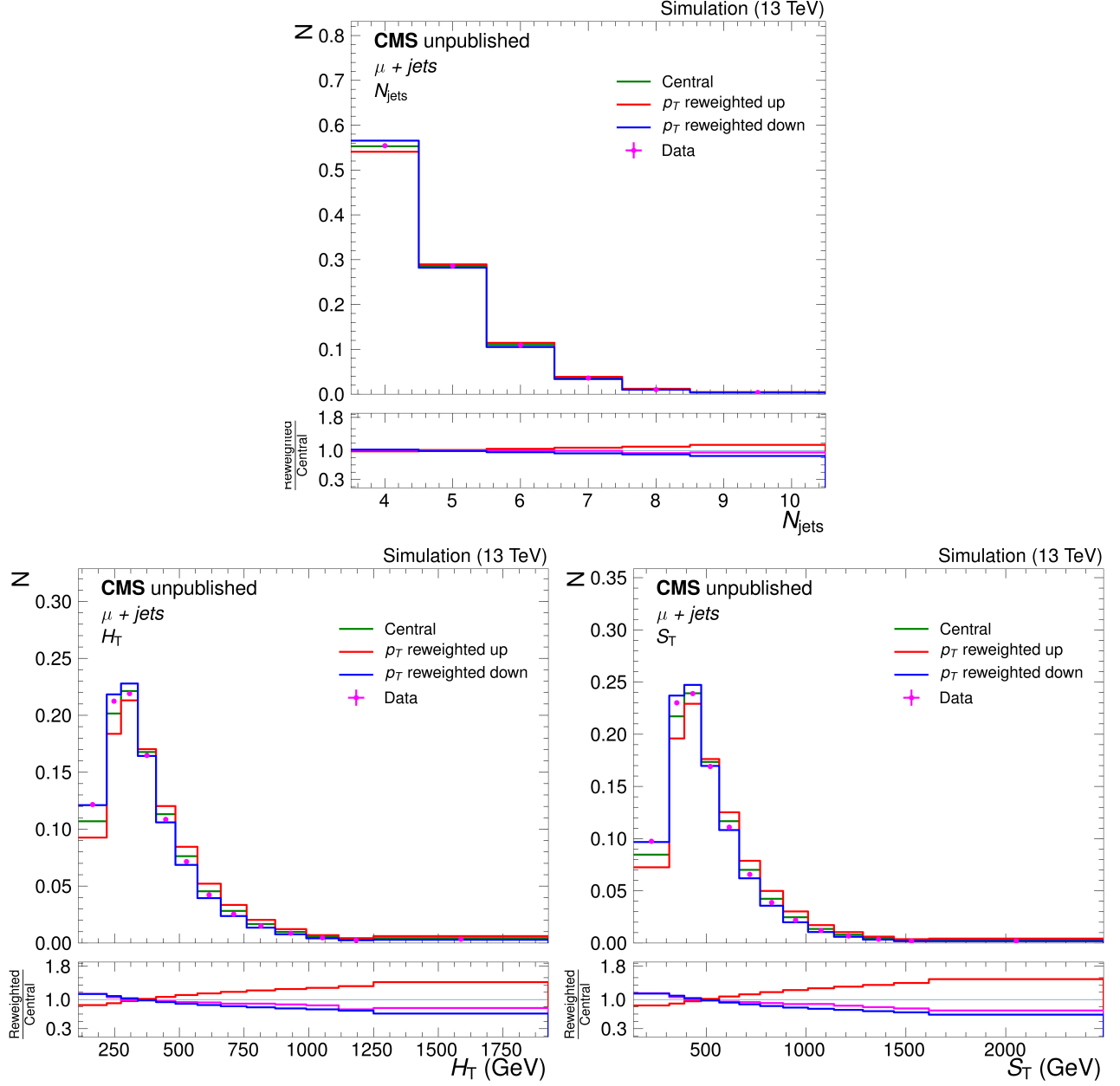


Figure 109: The kinematic event distributions given by the POWHEG+PYTHIA sample (green) with the top quark p_T reweighted up (red) and down (blue) to cover differences to data (magenta) in the $\mu + \text{jets}$ channel, for the N_{jets} , H_T and S_T event variables. The distributions are normalised to one.

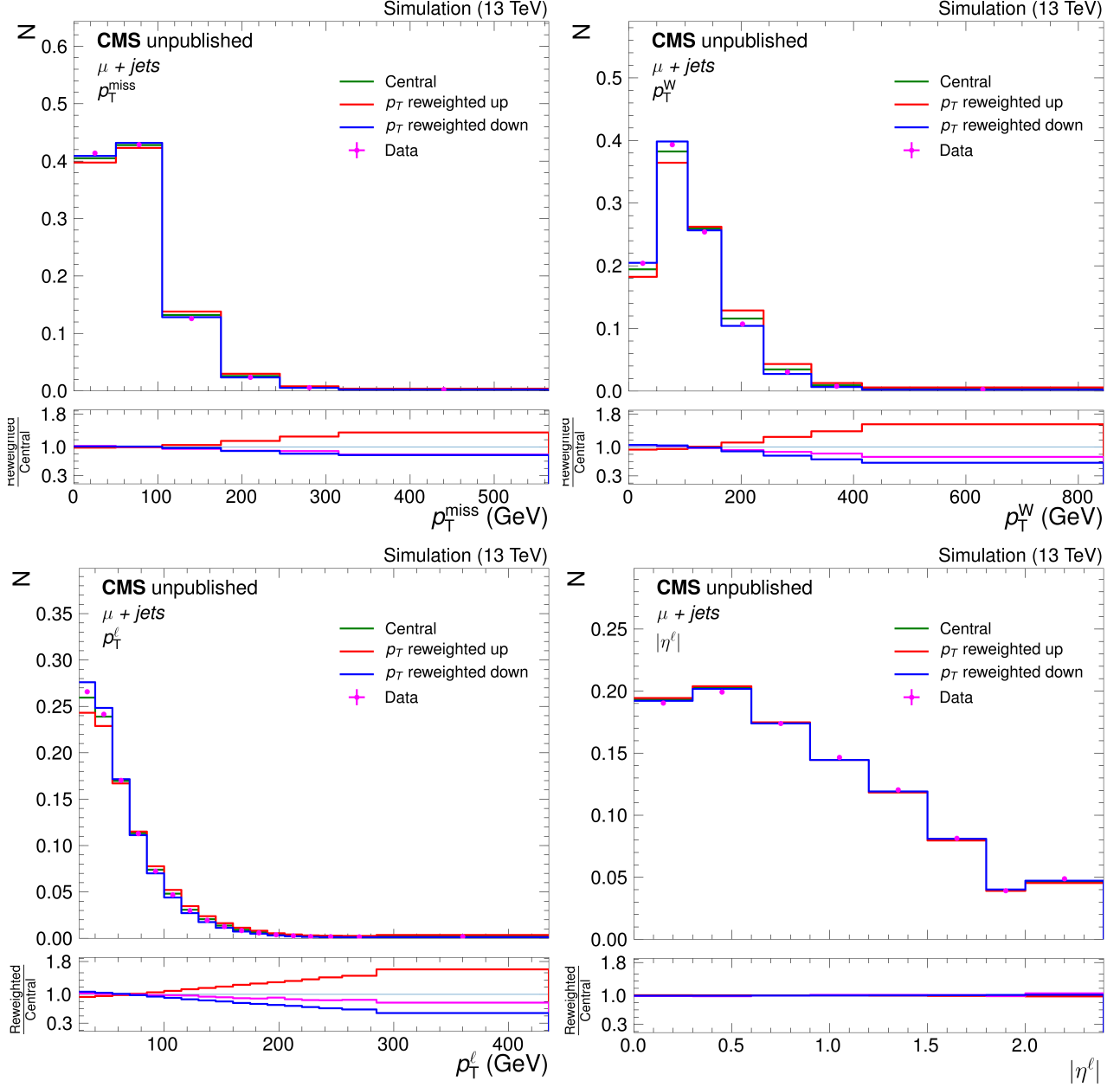


Figure 110: The kinematic event distributions given by the POWHEG+PYTHIA sample (green) with the top quark p_T reweighted up (red) and down (blue) to cover differences to data (magenta) in the $\mu + \text{jets}$ channel for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables. The distributions are normalised to one.

APPENDIX G

Checking bias in the muon decay channel and in alternate models

Bias tests performed in the μ +jets channel are shown in Figs. 111, 112. Additional bias tests on the unfolding procedure are performed by unfolding $t\bar{t}$ events generated from an alternative model and comparing to the true values. These are shown in Figs. 113, 114, 115 and 116 in the e+jets and μ +jets channels respectively. For the most part, they lie comfortably within the systematic uncertainties, shown as the grey band in the ratio plots.

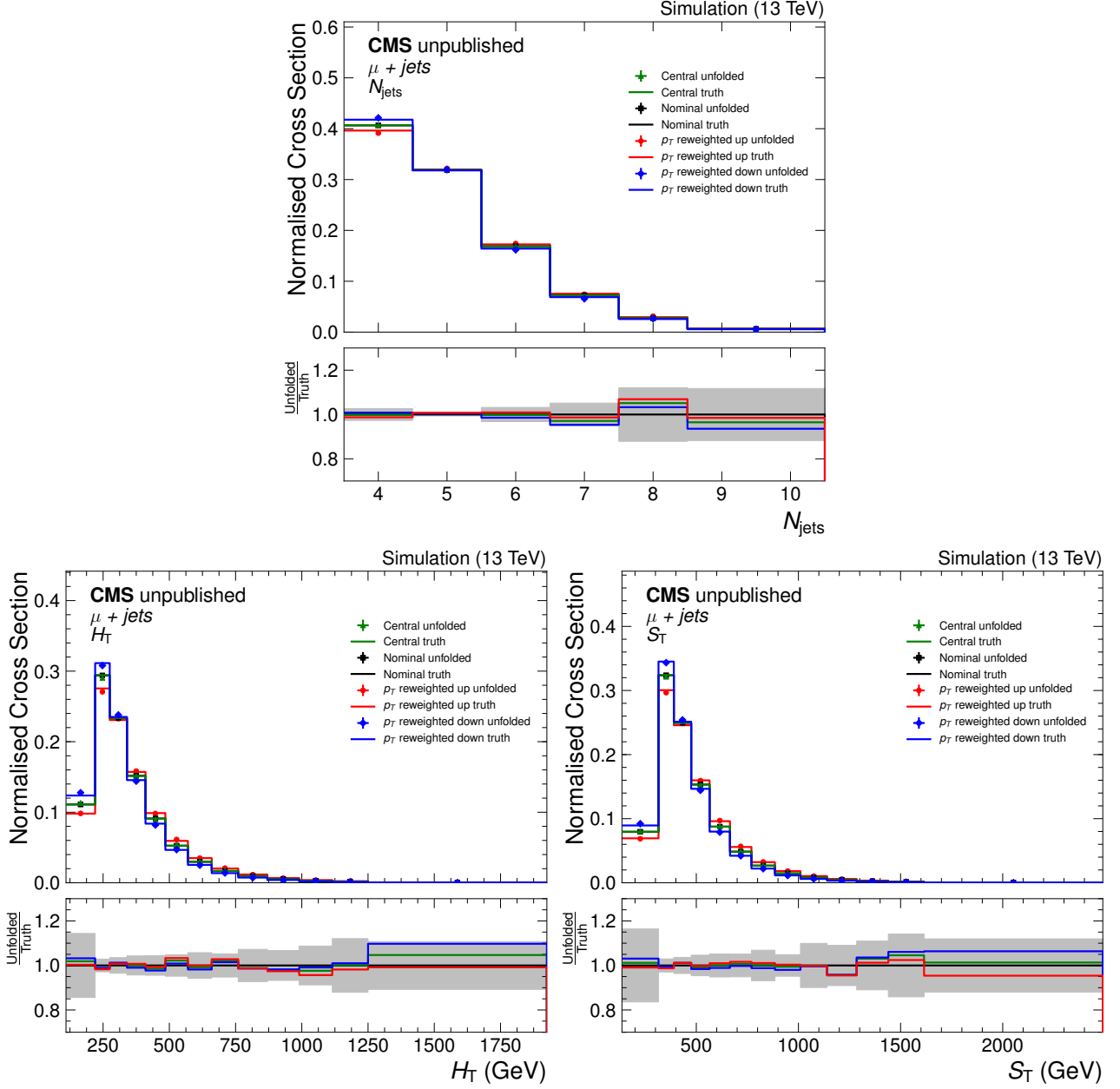


Figure 111: The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the $\mu + \text{jets}$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

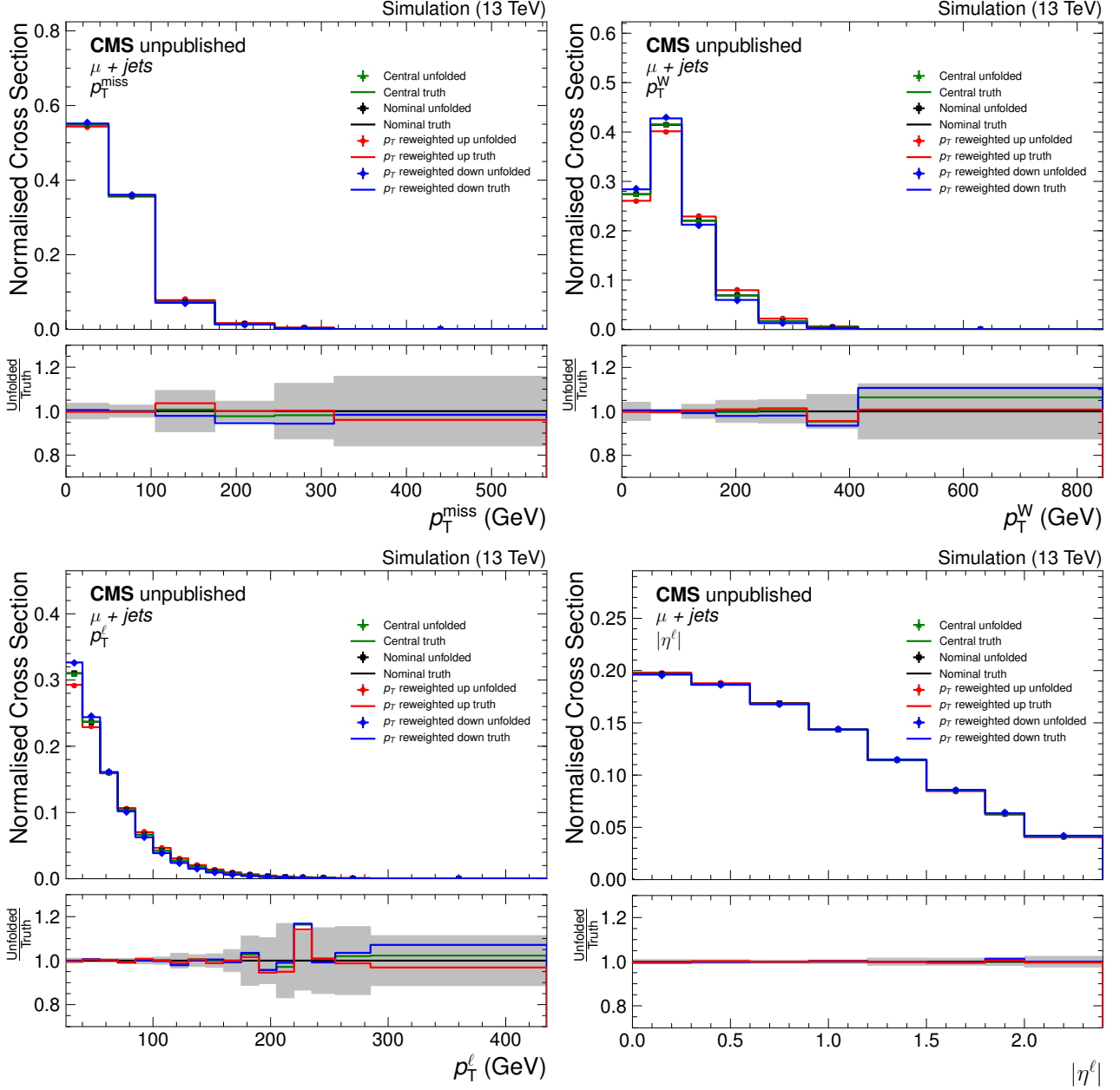


Figure 112: The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^{ℓ} and $|\eta^{\ell}|$ event variables in the $\mu + \text{jets}$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

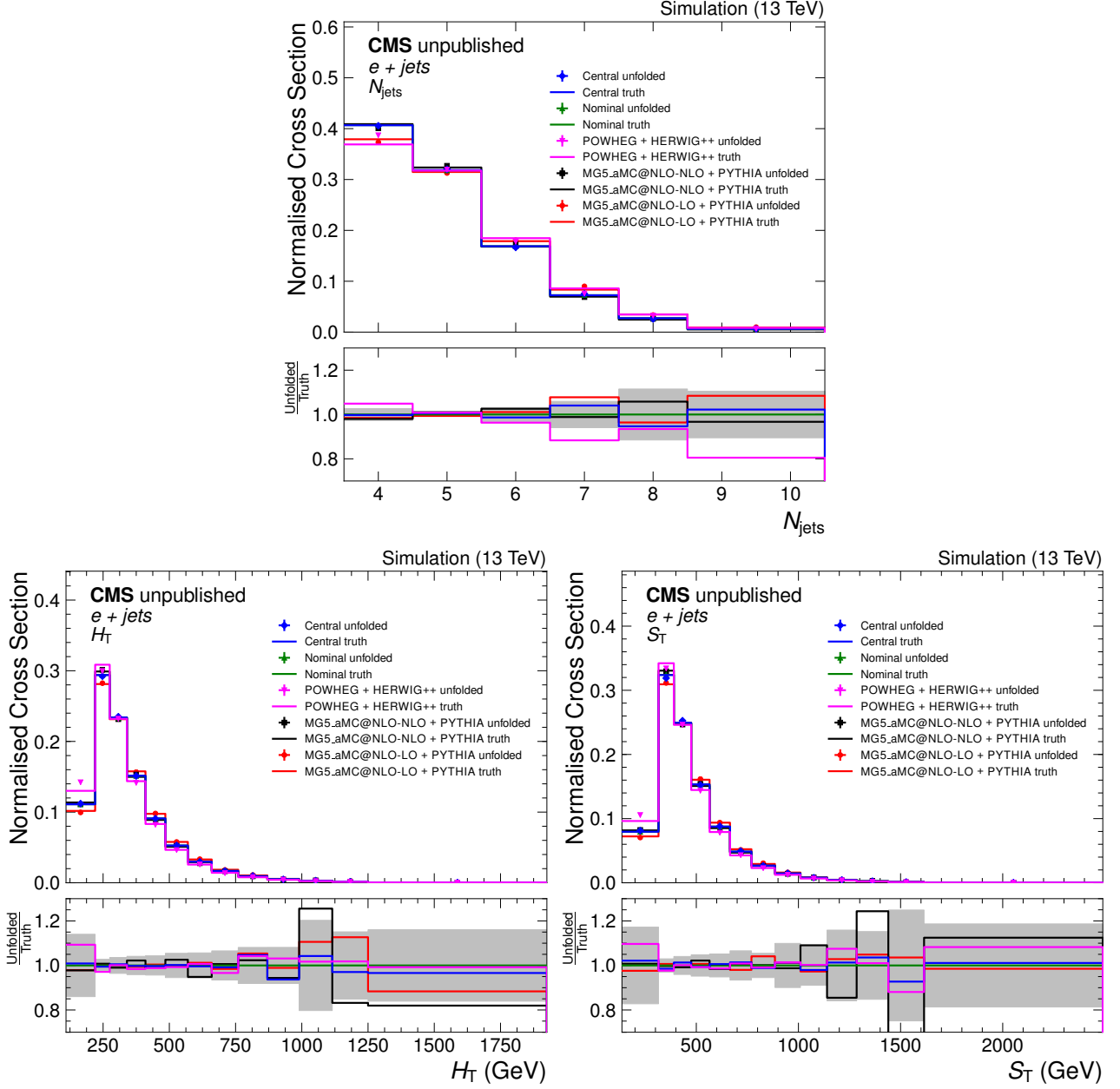


Figure 113: The cross sections for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the $e + \text{jets}$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

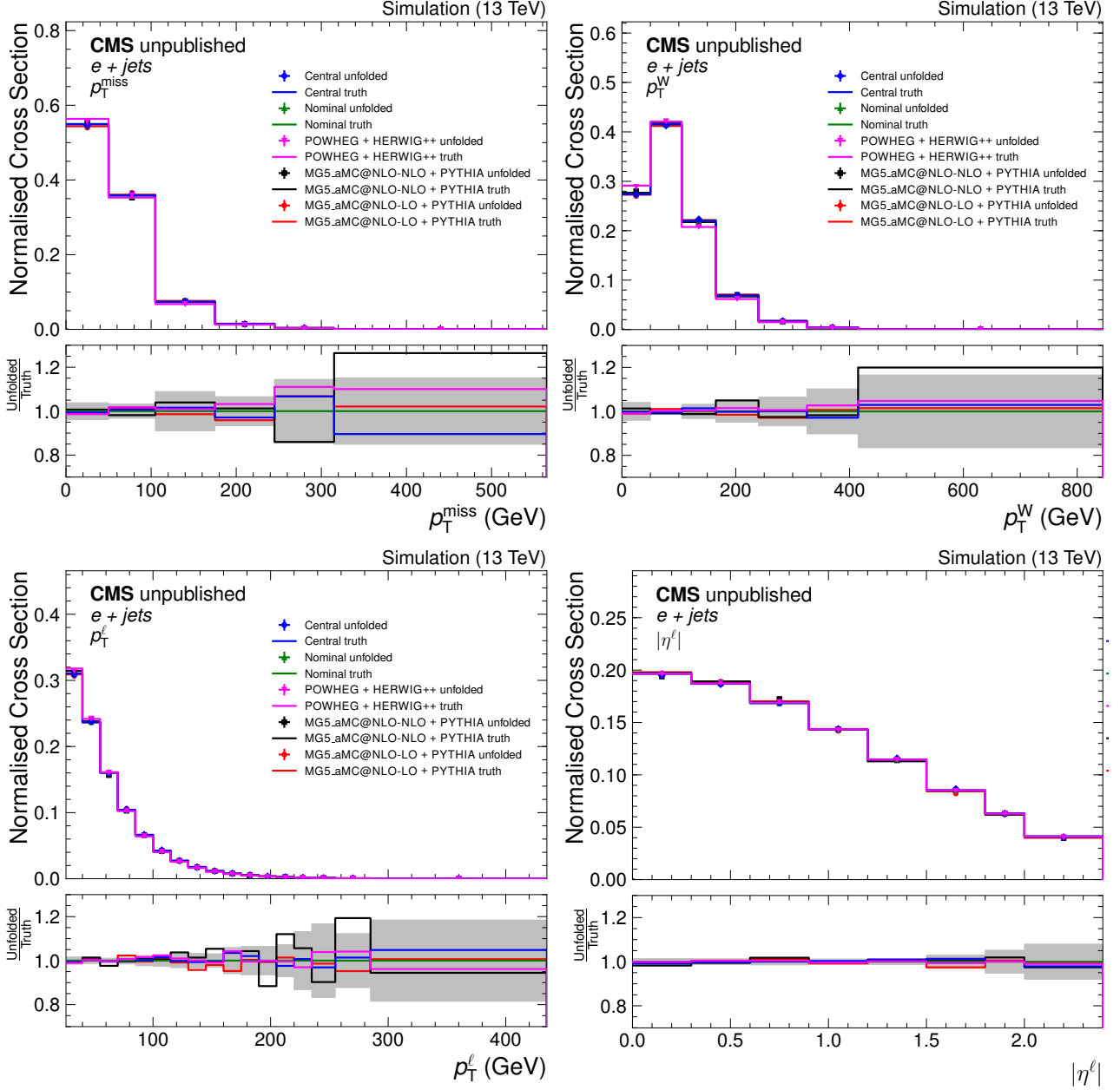


Figure 114: The cross sections for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^{ℓ} and $|\eta^{\ell}|$ event variables in the $e + jets$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

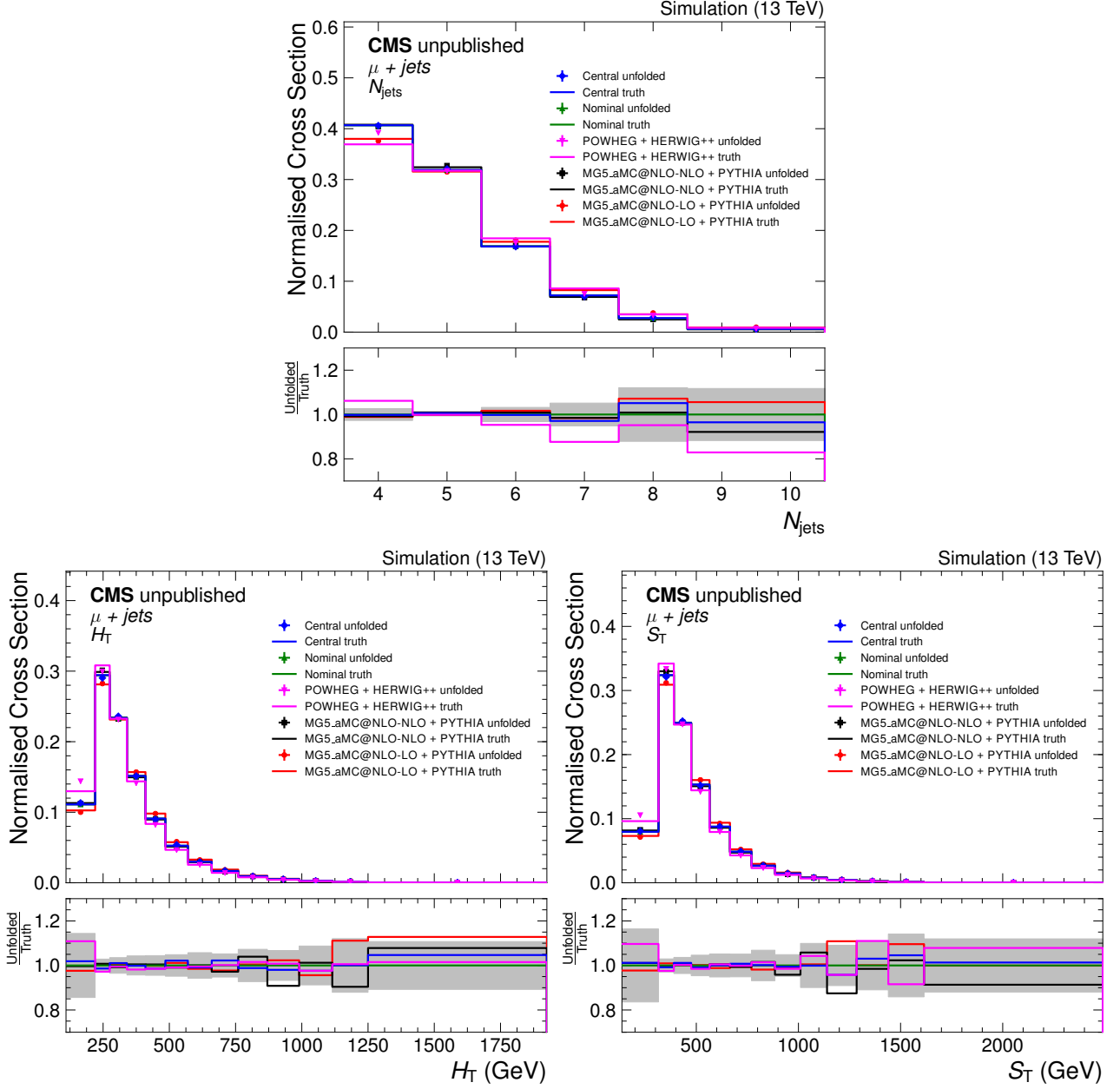


Figure 115: The cross sections for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the N_{jets} , H_T and S_T event variables in the $\mu + \text{jets}$ channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

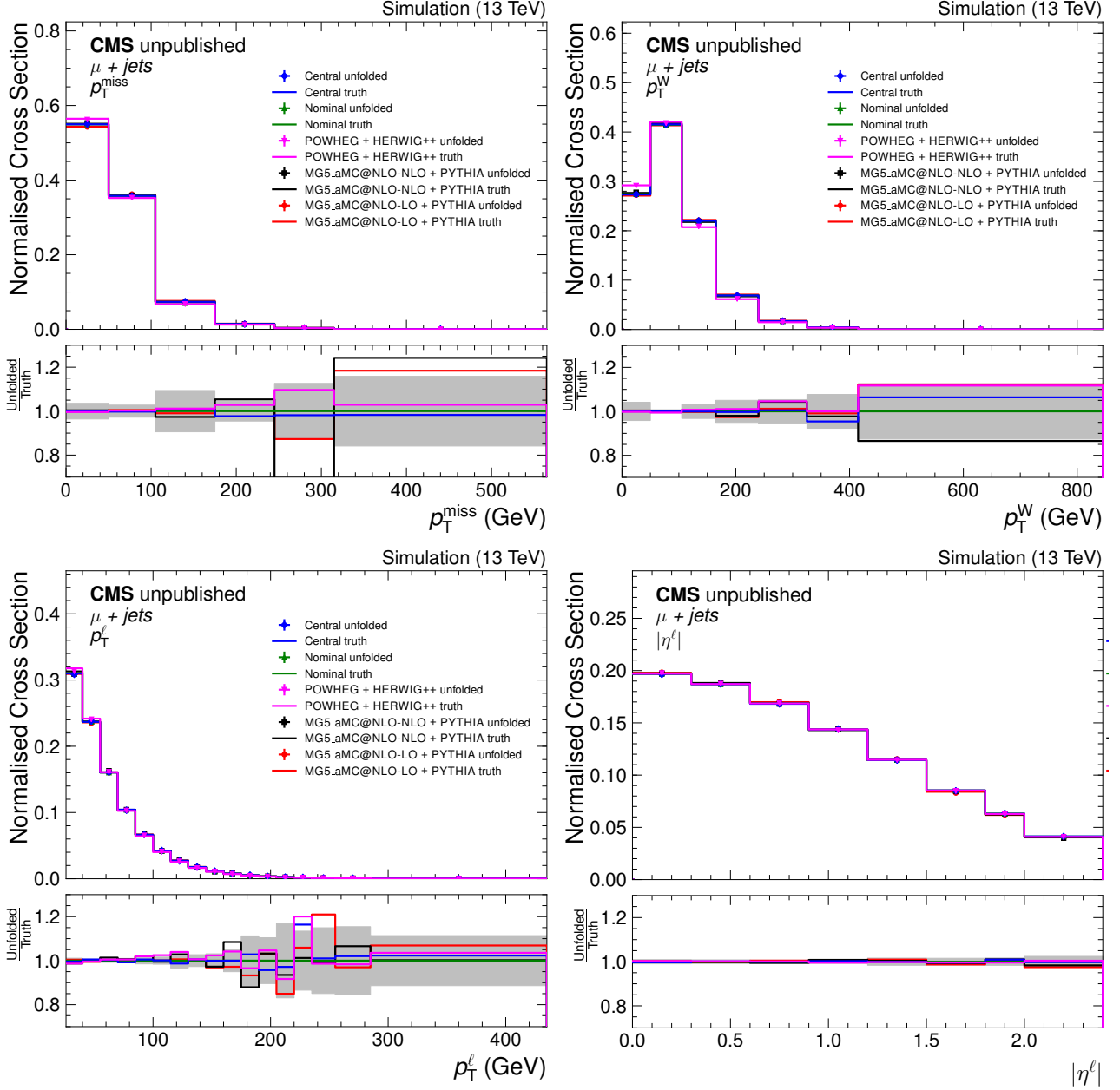


Figure 116: The cross sections for the alternate $t\bar{t}$ production models unfolded using the POWHEG+PYTHIA derived migration matrix compared to the true model cross sections are shown for the p_T^{miss} , p_T^W , p_T^ℓ and $|\eta^\ell|$ event variables in the μ +jets channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

APPENDIX H

Largest systematic uncertainties

Figures 117 and 118 portray the relative systematic uncertainties in every bin for each variable for the normalised cross section measurements. Similarly, Figs. 119 and 120 show the uncertainty compositions for the absolute cross section measurements. The gold band indicates the total systematic uncertainty and the grey band the total statistical uncertainty. Shown in bold are the systematic uncertainties which are the leading or subleading contributions to any bin.

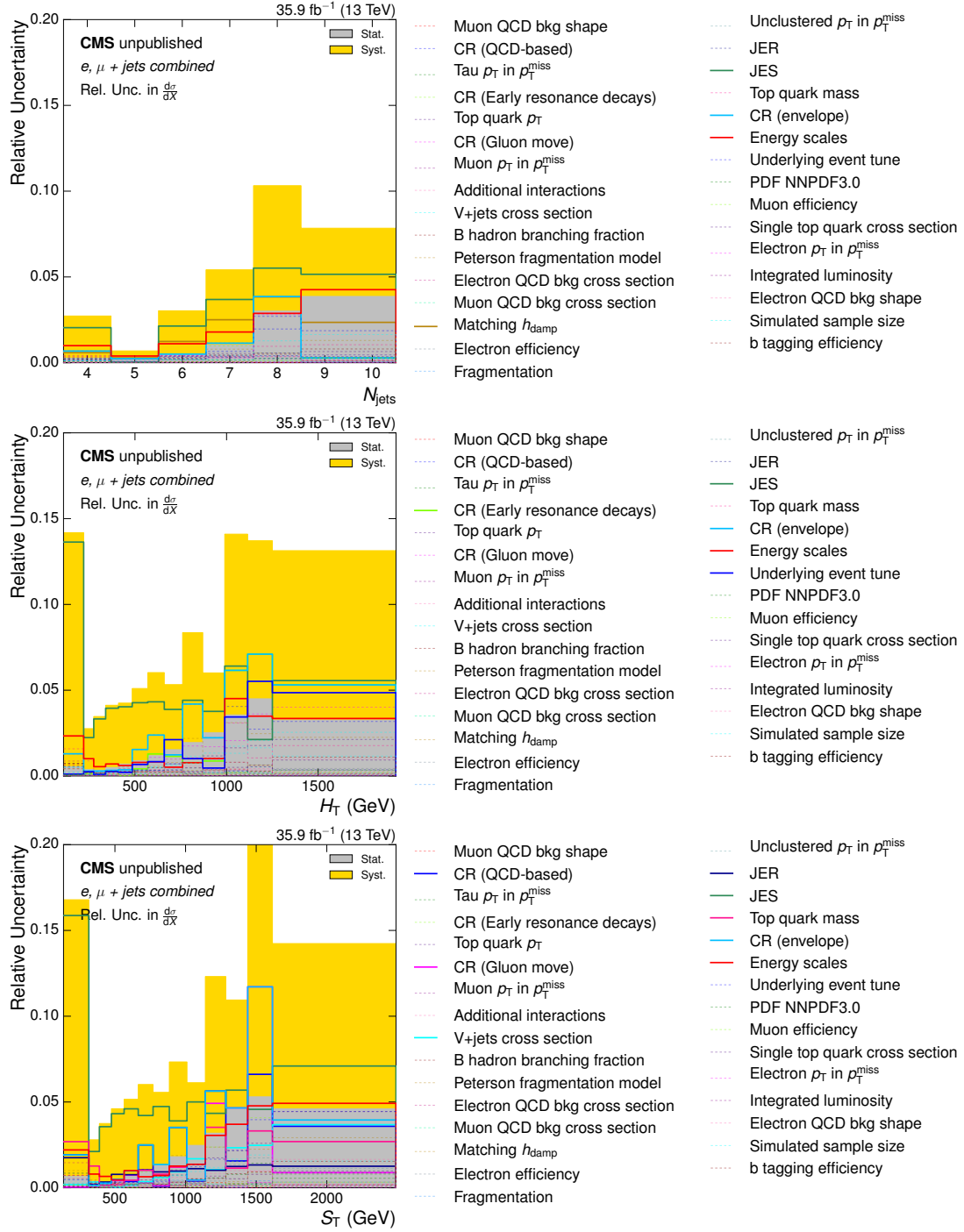


Figure 117: The composition of the systematic uncertainties for the N_{jets} , H_T and S_T event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

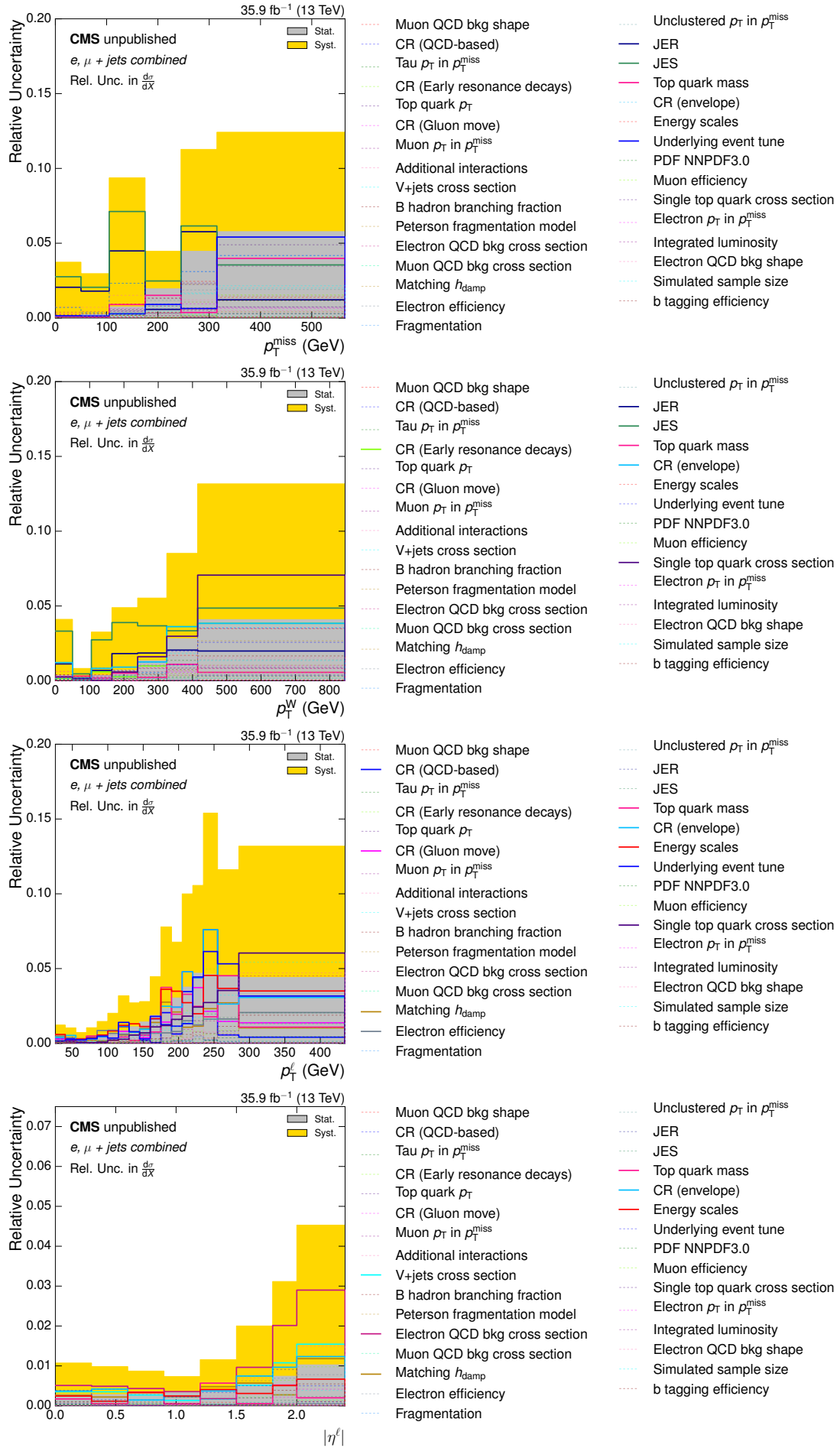


Figure 118: The composition of the systematic uncertainties for the p_T^{miss} , p_T^W , p_T^l and $|\eta^l|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

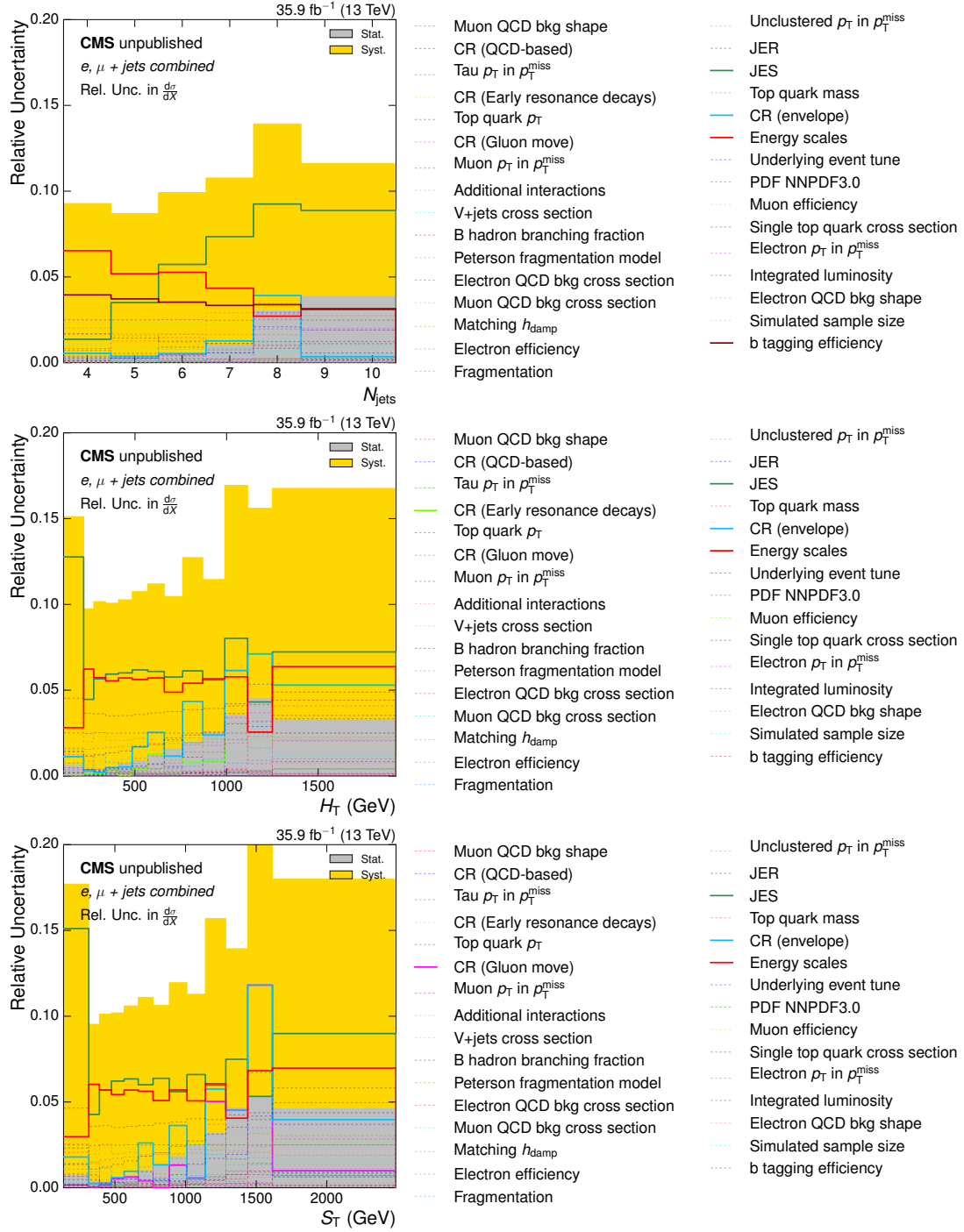


Figure 119: The composition of the systematic uncertainties for the N_{jets} , H_T and S_T event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

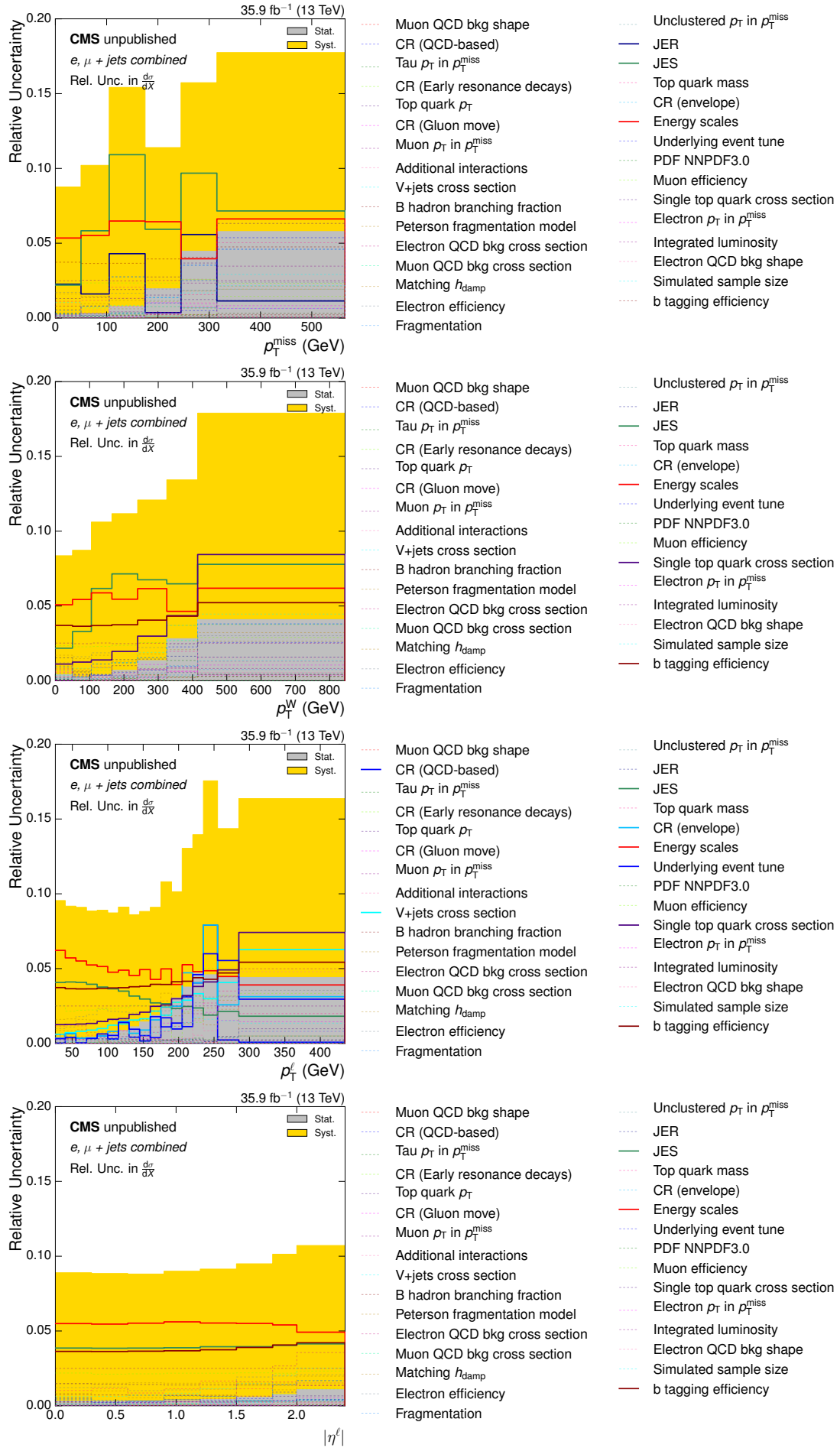


Figure 120: The composition of the systematic uncertainties for the p_T^{miss} , p_T^W , p_T^l and $|\eta^\ell|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

APPENDIX I

Covariance Matrices

Figures 121 and 122 show the total covariance matrices calculated for the normalised differential cross section measurement with respect to each kinematic event variable. The total covariance matrix is calculated from the sum of the covariances matrices from each source of uncertainty. The total covariances for the absolute differential cross sections are measured similarly and can be found in Figs. 123 and 124

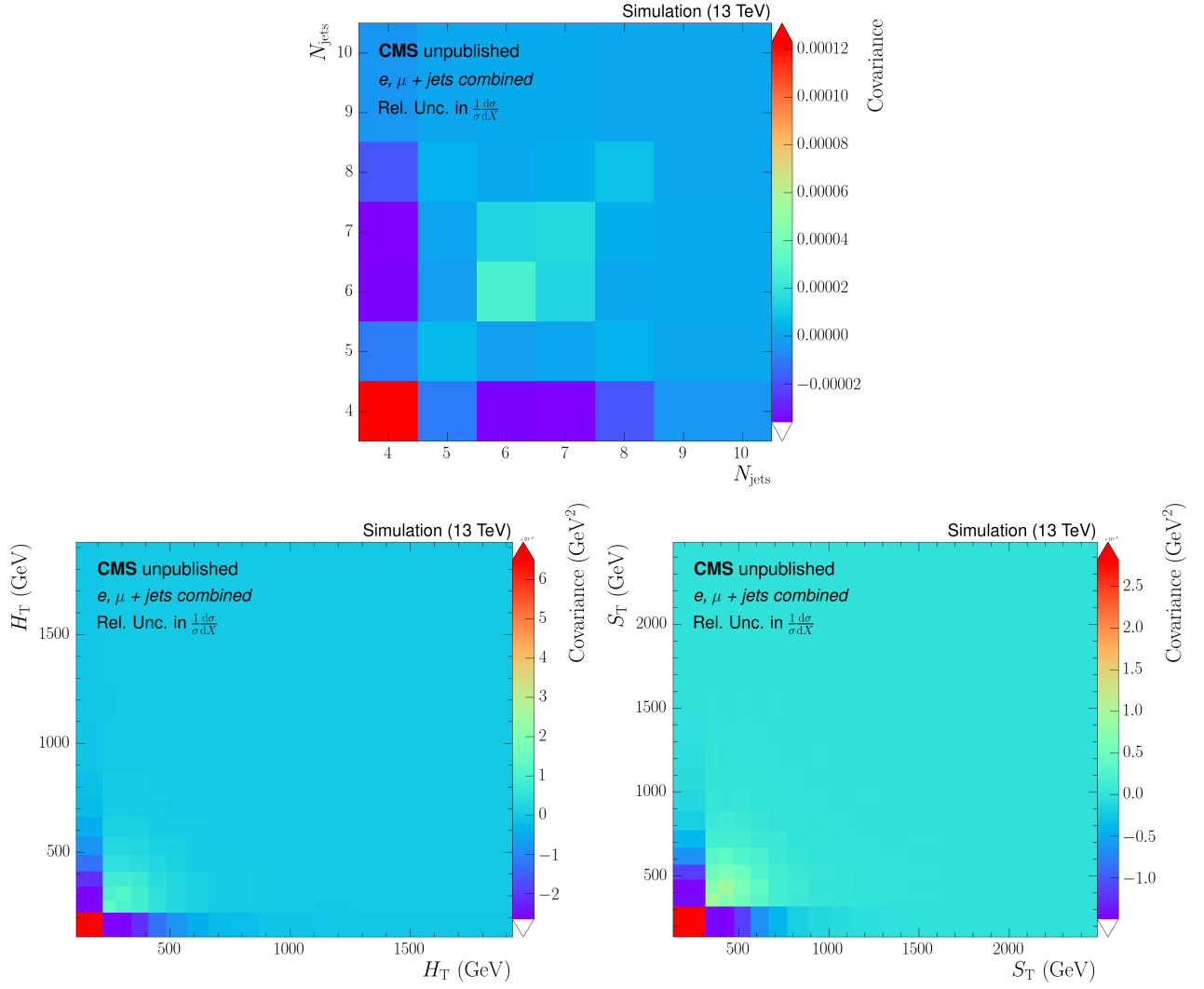


Figure 121: The total covariance matrix for the normalised differential cross section measurements with respect to the N_{jets} , H_{T} , and S_{T} event variables.

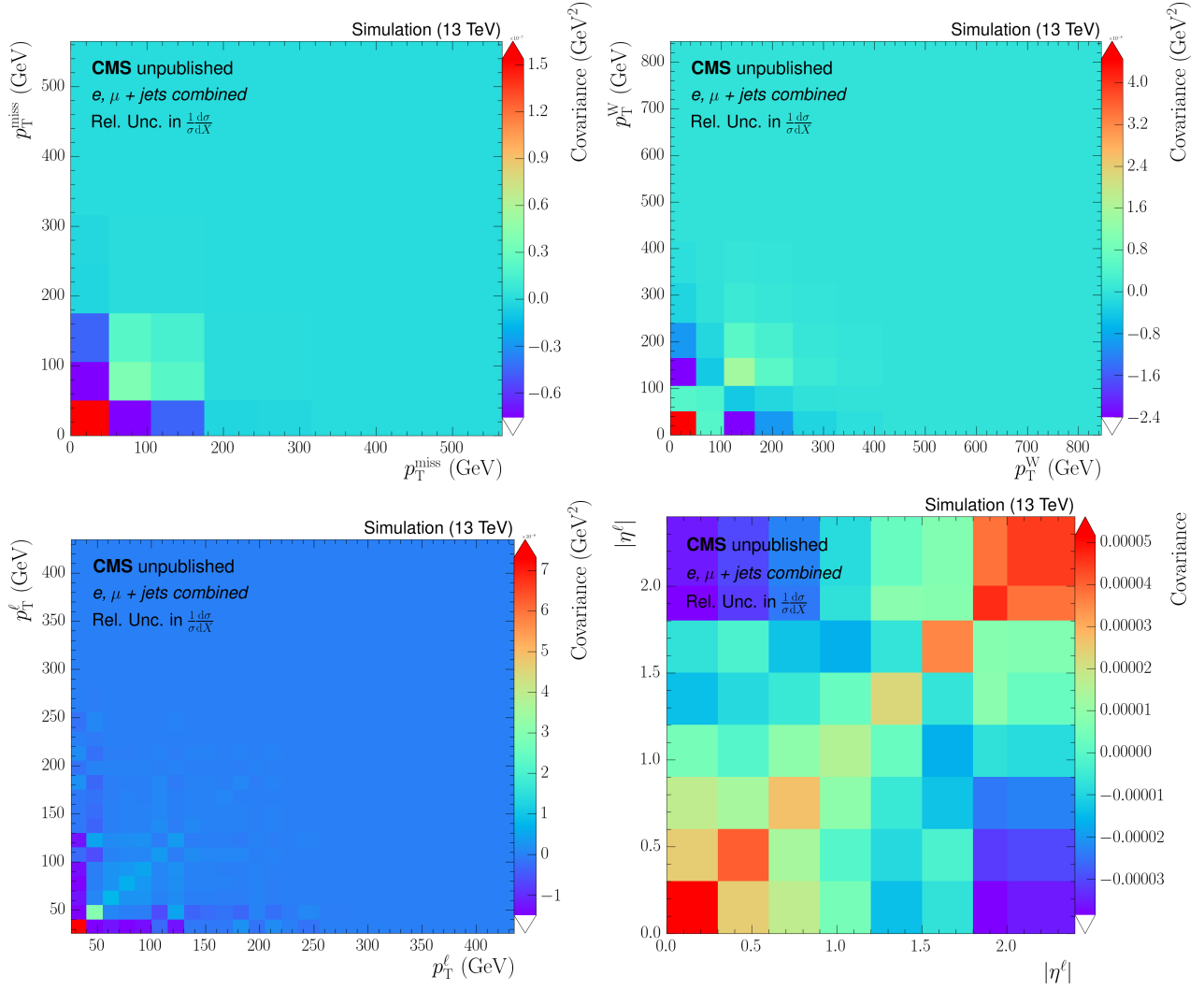


Figure 122: The total covariance matrix for the normalised differential cross section measurements with respect to the p_T^{miss} , p_T^{W} , p_T^{ℓ} and $|\eta^{\ell}|$ event variables.

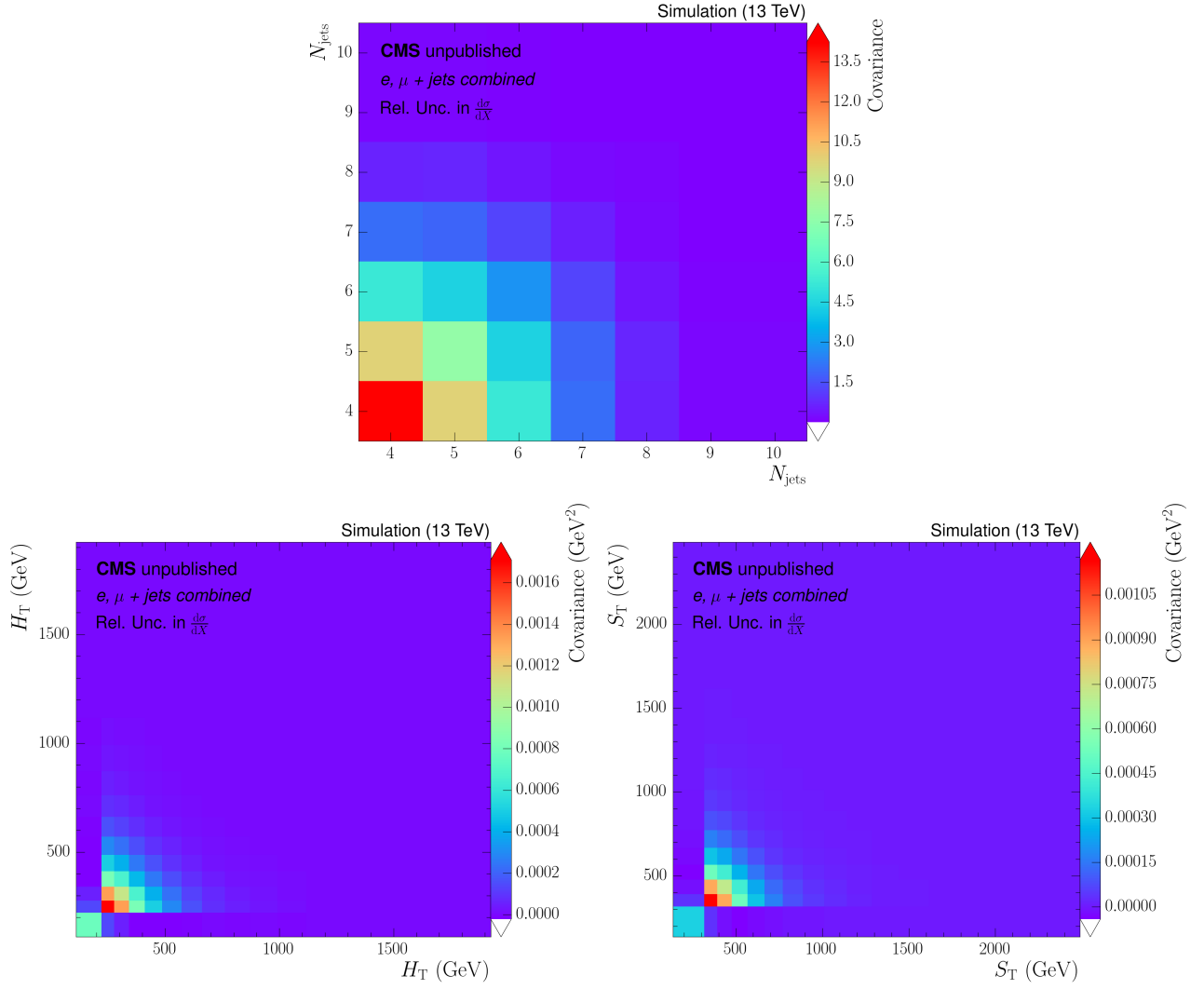


Figure 123: The total covariance matrix for the absolute differential cross section measurements with respect to the N_{jets} , H_T , and S_T event variables.

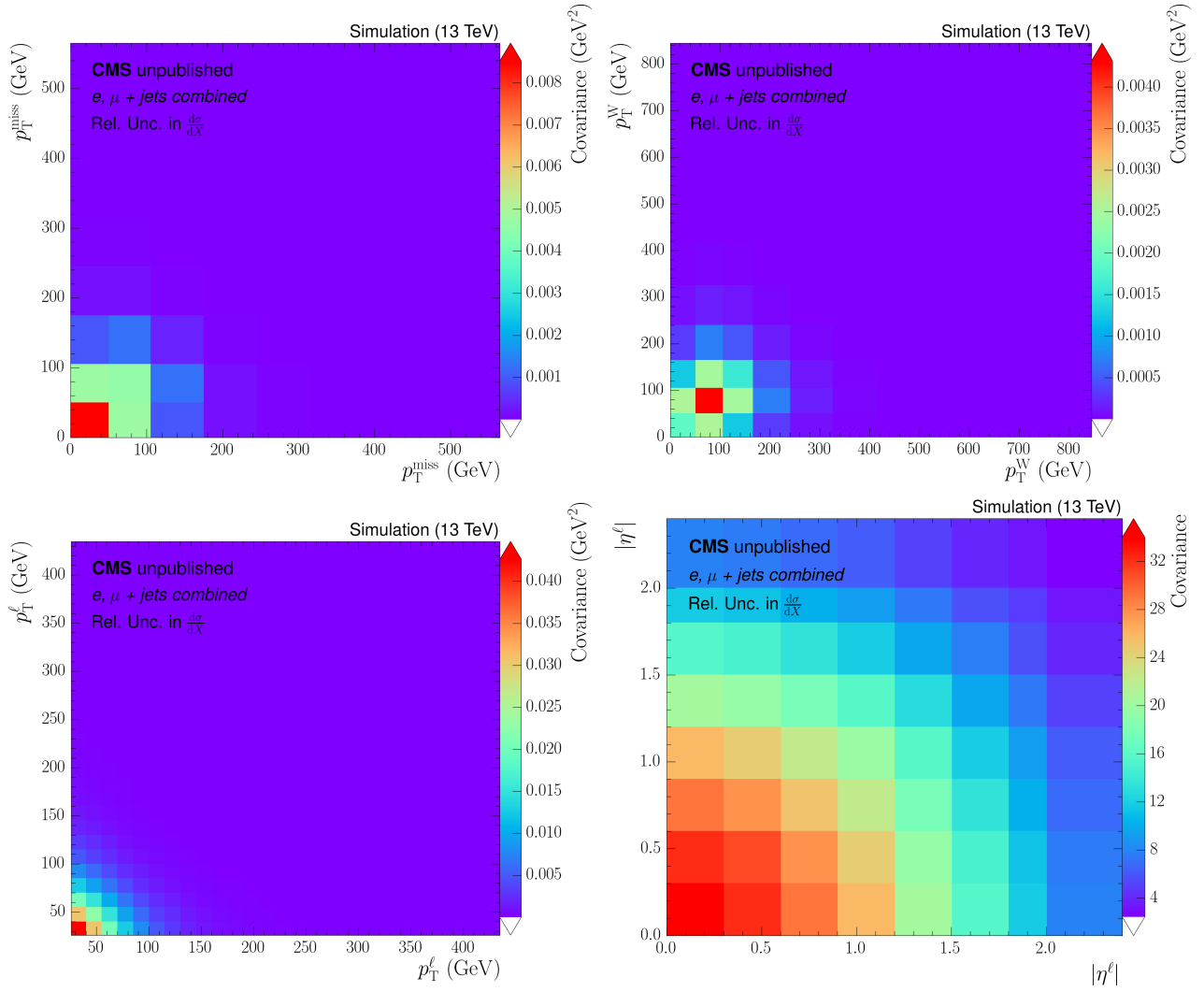


Figure 124: The total covariance matrix for the absolute differential cross section measurements with respect to the p_T^{miss} , p_T^{W} , p_T^{ℓ} and $|\eta^{\ell}|$ event variables.

APPENDIX J

Differential cross sections as a function of lepton pseudorapidity

As has been seen, there is some tension in the differential cross sections calculated as a function of $|\eta^\ell|$. Tables 19 and 20 along with Fig. 125 show the χ^2 goodness-of-fit tests and differential cross section plots with respect to the $|\eta^\ell|$ event variable in the e+jets and μ +jets channels. This tension in the combined differential cross section measurements seems to come from the muon channel. It is likely that it is caused by an effect on the shape of the distribution, possibly introduced when scale factors have been applied.

Table 19: Results of a goodness-of-fit test between the normalised cross sections as a function of $|\eta^\ell|$ in data and several models, for the e+jets, μ +jets and e, μ +jets channels. The values are given as χ^2 /number of degrees of freedom (ndf).

$ \eta^\ell $ Normalised	POWHEG+PYTHIA		With MC Theory		POWHEG+HERWIG++		MG5_aMC@NLO-NLO		MG5_aMC@NLO-LO	
	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value
e, μ +jets	19 / 7	< 0.01	15 / 7	0.04	19 / 7	< 0.01	24 / 7	< 0.01	30 / 7	< 0.01
e+jets	6.5 / 7	0.49	5.1 / 7	0.64	4.7 / 7	0.69	6.8 / 7	0.45	7.5 / 7	0.38
μ +jets	17 / 7	0.02	11 / 7	0.14	18 / 7	0.01	25 / 7	< 0.01	24 / 7	< 0.01

Table 20: Results of a goodness-of-fit test between the absolute cross sections as a function of $|\eta^\ell|$ in data and several models, for the e+jets, μ +jets and e, μ +jets channels. The values are given as χ^2 /number of degrees of freedom (ndf).

$ \eta^\ell $ Absolute	POWHEG+PYTHIA		With MC Theory		POWHEG+HERWIG++		MG5_aMC@NLO-NLO		MG5_aMC@NLO-LO	
	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value	χ^2 /ndf	p-value
e, μ +jets	16 / 8	0.04	15 / 8	0.06	16 / 8	0.04	23 / 8	< 0.01	31 / 8	< 0.01
e+jets	7.4 / 8	0.50	7.1 / 8	0.52	6.7 / 8	0.57	9.0 / 8	0.34	12 / 8	0.14
μ +jets	16 / 8	0.05	12 / 8	0.17	16 / 8	0.04	25 / 8	< 0.01	27 / 8	< 0.01

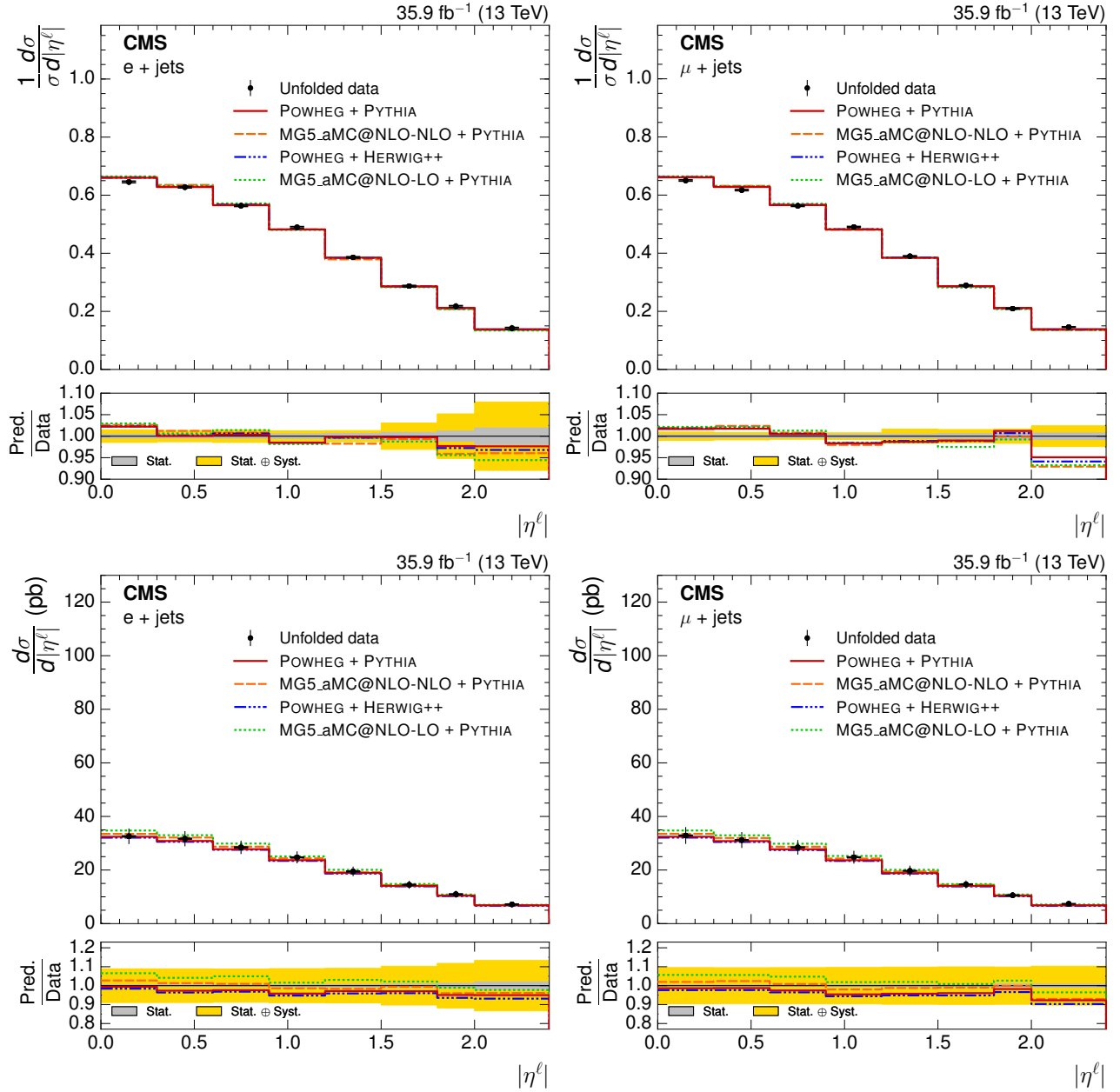


Figure 125: The normalised (upper) and absolute (lower) differential cross sections with respect to the $|\eta^\ell|$ event variable, in the e+jets (left) and μ +jets (right) channels.

APPENDIX K

Additional double differential distributions

Figures 126, 127 and 128 show the basic control plots, response matrices and double differential cross section measurements for the (S_T, p_T^{miss}) set of event variables. This gives a comparison for how well the total activity of a $t\bar{t}$ event is modelled depending on the missing p_T .

Another set of event variables, $(p_T^{\text{miss}}, N_{\text{jets}})$, shows how well the missing p_T is modelled for different jet multiplicities and is shown in Figs. 129, 130 and 131.

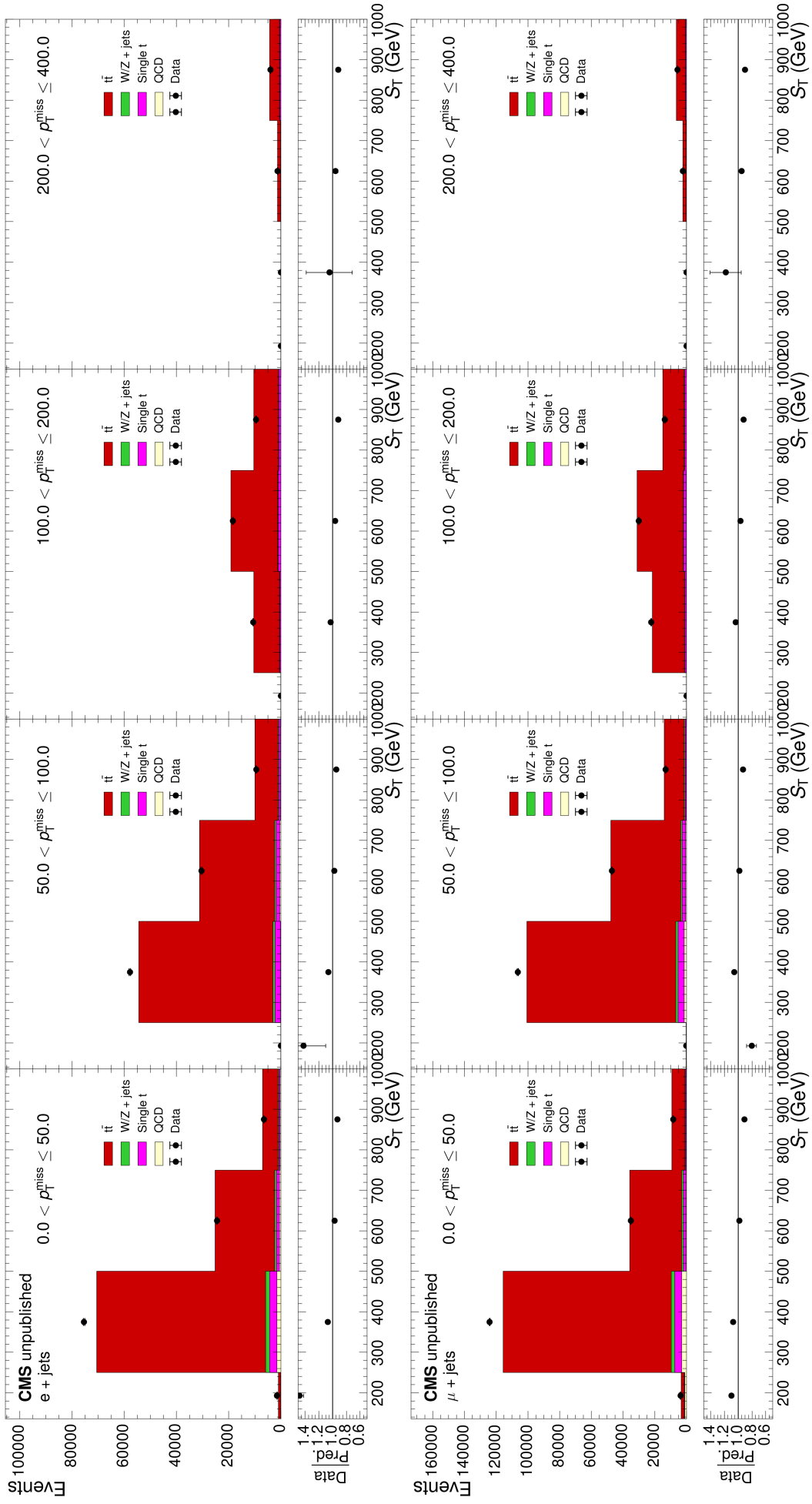


Figure 126: The $t\bar{t}$ yield after full event selection in the $e+\text{jets}$ (upper) and $\mu+\text{jets}$ (lower) channels as a function of the S_T and p_T^{miss} variables. Only statistical uncertainties are shown.

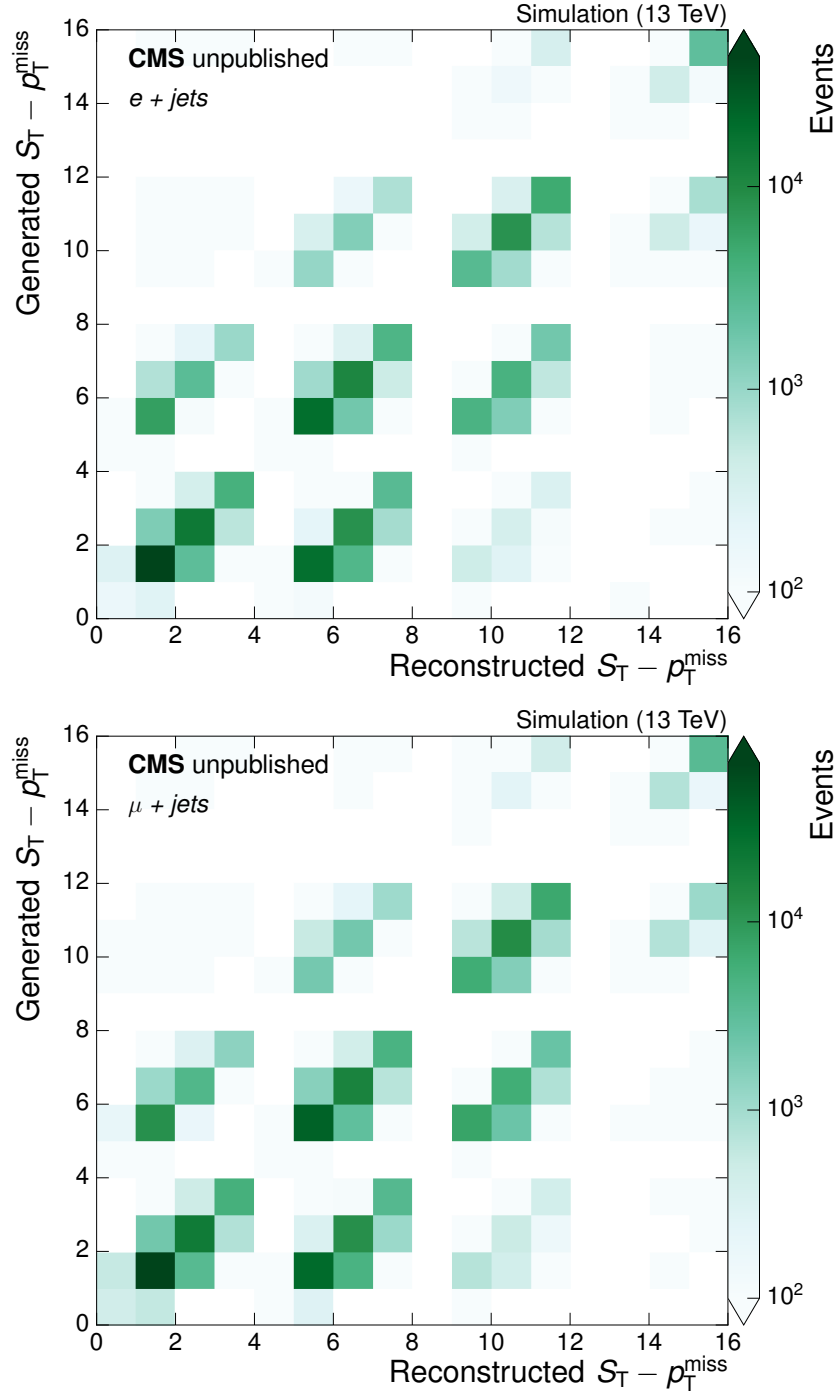


Figure 127: The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the S_T and p_T^{miss} events variables, in the $e + \text{jets}$ (upper) and $\mu + \text{jets}$ (lower) channels.

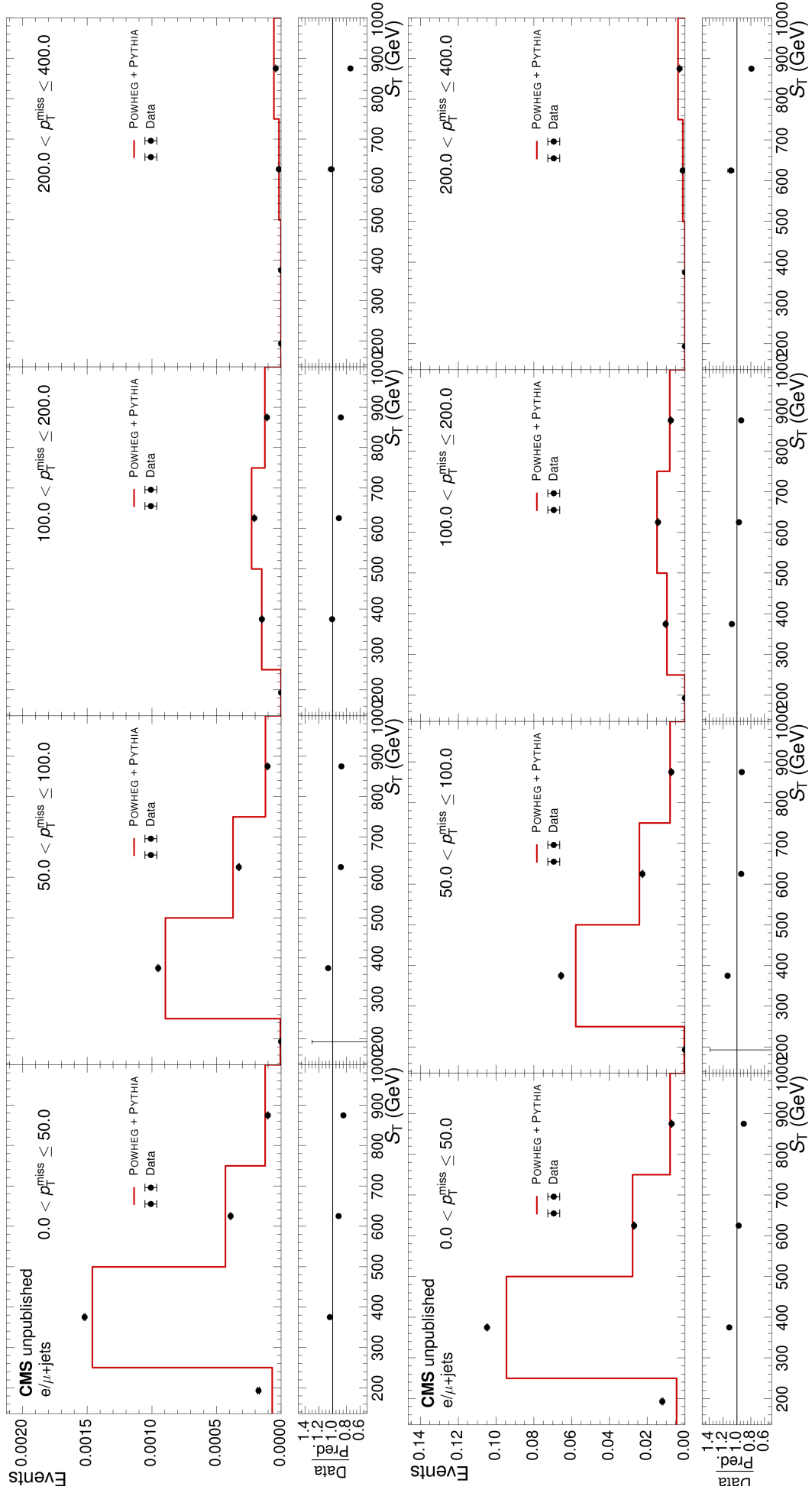


Figure 128: Normalised (upper) and absolute (lower) double differential $t\bar{t}$ production cross section measurements with respect to the S_T and p_T^{miss} event variables in comparison to the POWHEG+PYTHIA model. Only statistical uncertainties are shown.

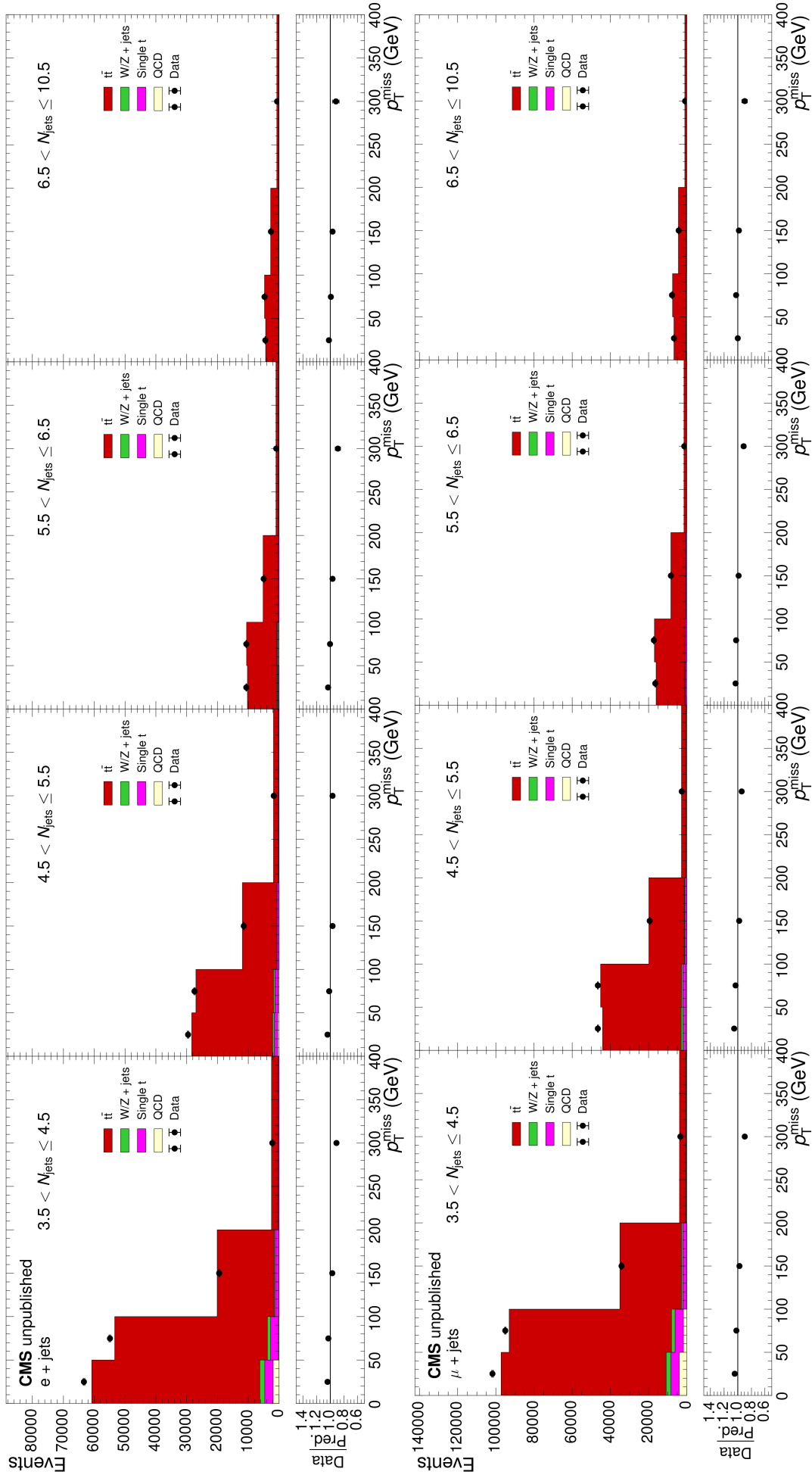


Figure 129: The $t\bar{t}$ yield after full event selection in the $e+jets$ (upper) and $\mu+jets$ (lower) channels as a function of the p_T^{miss} and N_{jets} variables. Only statistical uncertainties are shown.

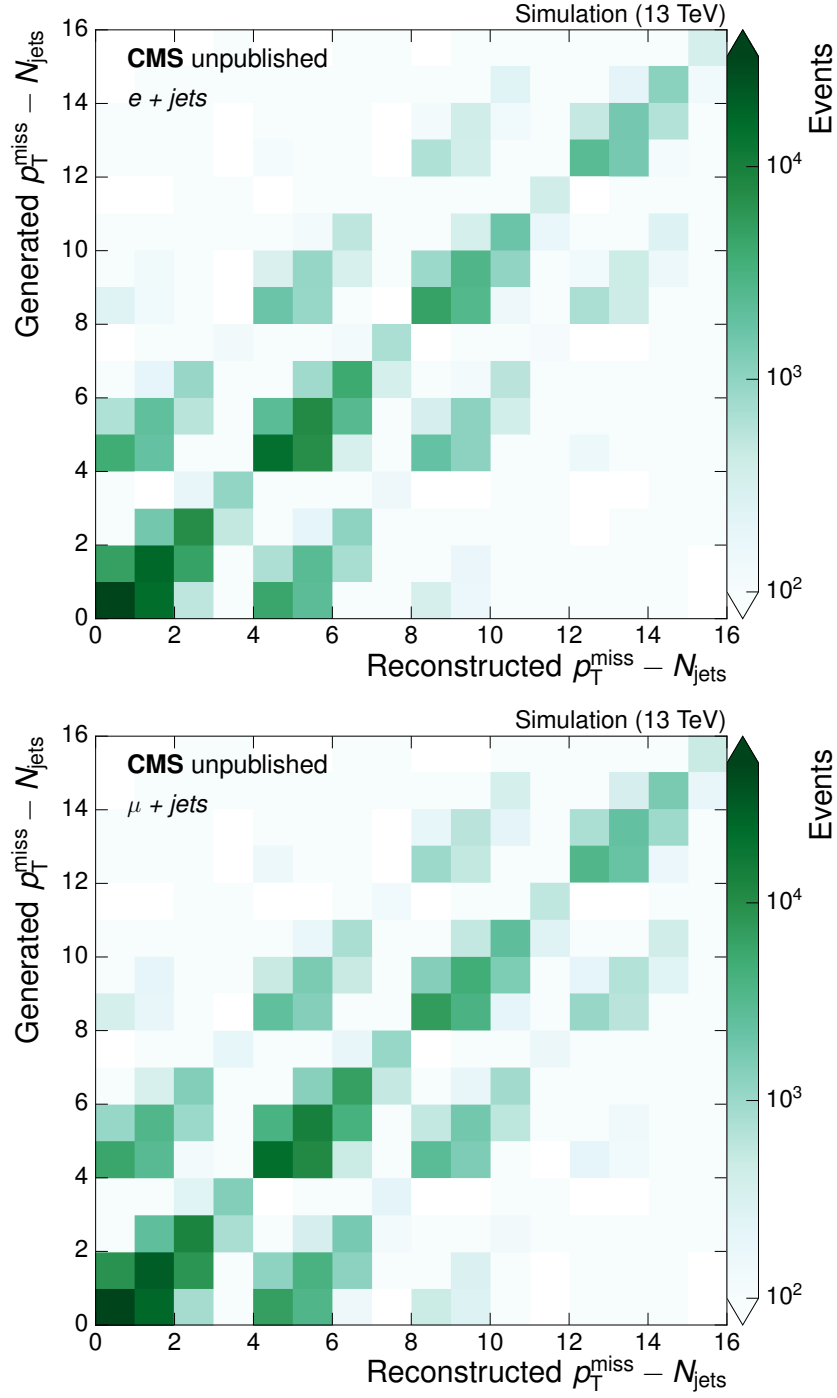


Figure 130: The migration matrices constructed, as a function of bin number, for the double differential cross section measurements with respect to the p_T^{miss} and N_{jets} events variables, in the $e + jets$ (upper) and $\mu + jets$ (lower) channels.

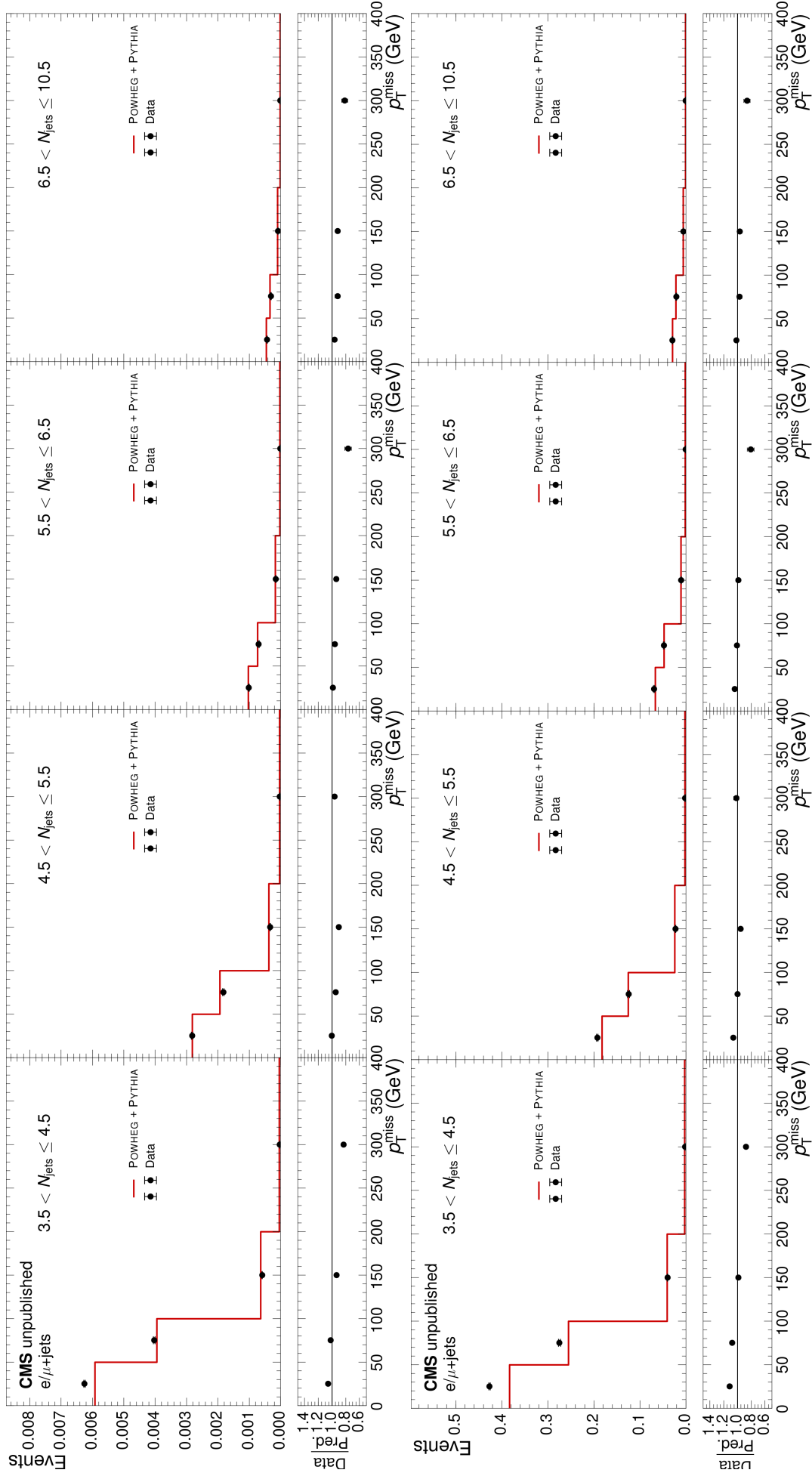


Figure 131: Normalised (upper) and absolute (lower) double differential $t\bar{t}$ production cross section measurements with respect to the p_T^{miss} and N_{jets} event variables in comparison to the POWHEG+PYTHIA model. Only statistical uncertainties are shown.

APPENDIX L

EFT studies

Figures 132, 133, 134, 135, 136 and 137 show the predictions using a LO EFT theory model simulated with MG5_aMC@NLO-LO and PYTHIA compared to the absolute differential $t\bar{t}$ cross sections with respect to N_{jets} , S_{T} , $p_{\text{T}}^{\text{miss}}$, p_{T}^{W} , p_{T}^{ℓ} and $|\eta^{\ell}|$ in the left panels and the corresponding nominal fit for the $\Delta\chi^2$ as a function of $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strength in the right panels. Also shown in the right panels are the best fit value for the $C_{\text{tG}}/\Lambda_{\text{NP}}^2$ strength together with the 68% and 95% confidence intervals. A summary of the 95% confidence intervals obtained with each event variable is shown in Tab. 21.

Table 21: The 95% confidence intervals reached for each event variable

Event Variable	95% CI
N_{jets}	$-1.73 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < -0.28$
H_{T}	$-1.86 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.29$
S_{T}	$-1.37 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.58$
$p_{\text{T}}^{\text{miss}}$	$-1.44 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.51$
p_{T}^{W}	$-1.65 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.20$
p_{T}^{ℓ}	$-1.95 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 0.21$
$ \eta^{\ell} $	$-1.27 < C_{\text{tG}}/\Lambda_{\text{NP}}^2 < 1.14$

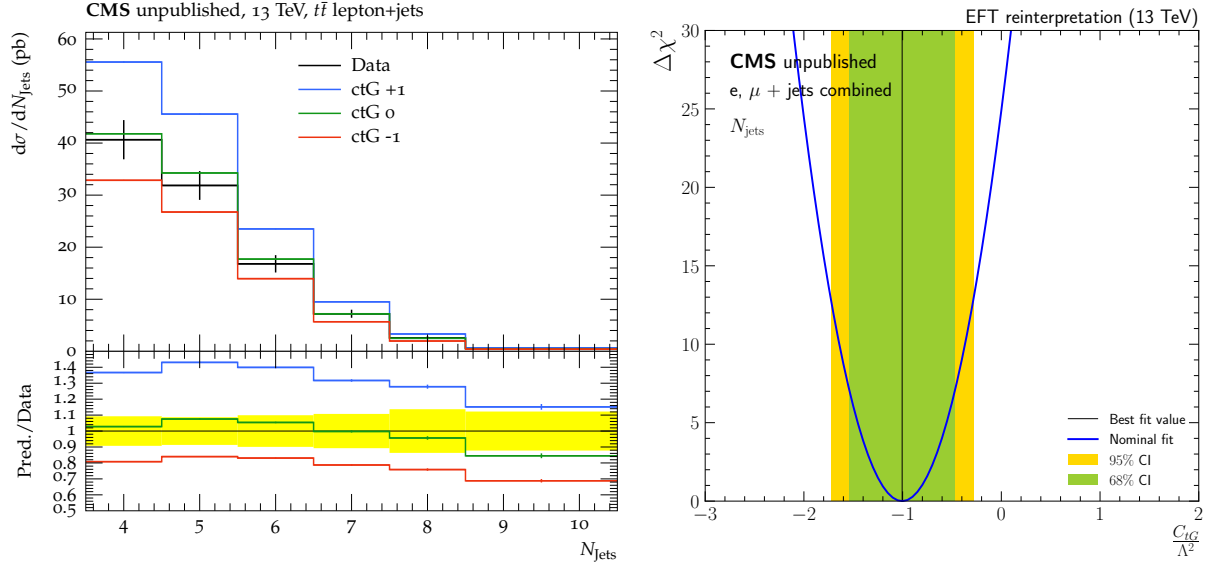


Figure 132: The left panel shows the predictions of EFT simulations using $C_{tG}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the N_{jets} event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{tG}/\Lambda_{\text{NP}}^2$ strengths.

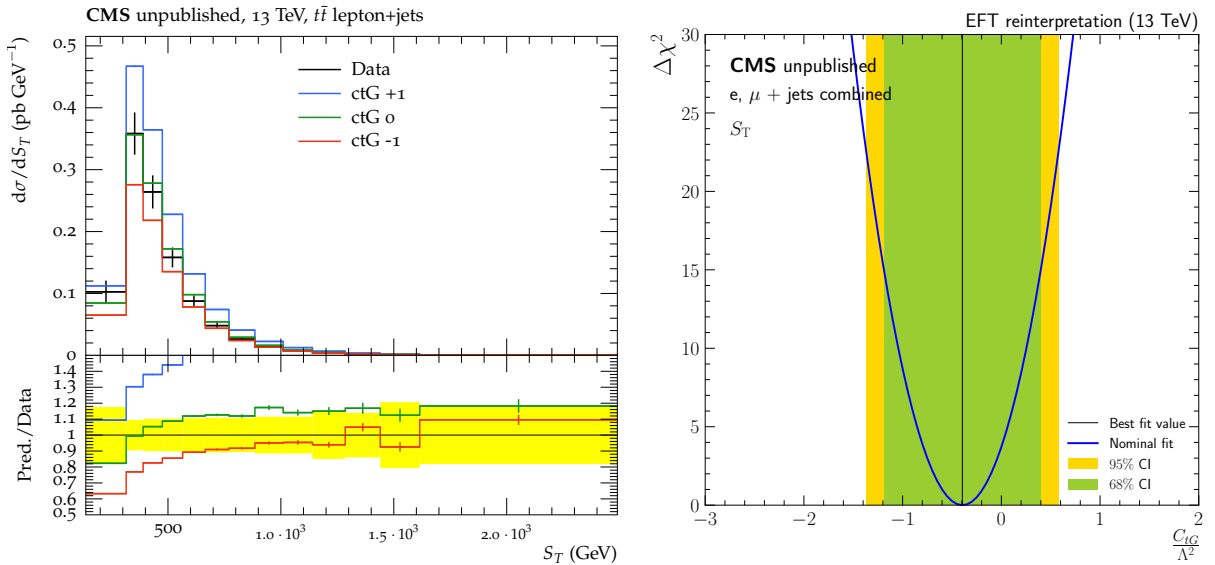


Figure 133: The left panel shows the predictions of EFT simulations using $C_{tG}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the S_T event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{tG}/\Lambda_{\text{NP}}^2$ strengths.

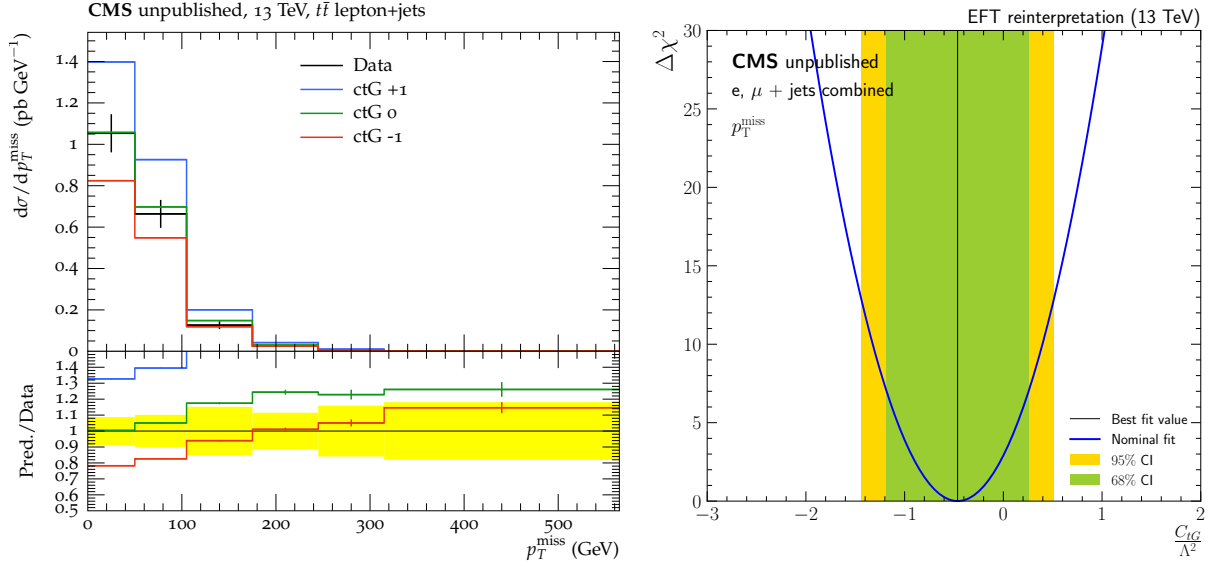


Figure 134: The left panel shows the predictions of EFT simulations using $C_{tG}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the p_T^{miss} event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{tG}/\Lambda_{\text{NP}}^2$ strengths.

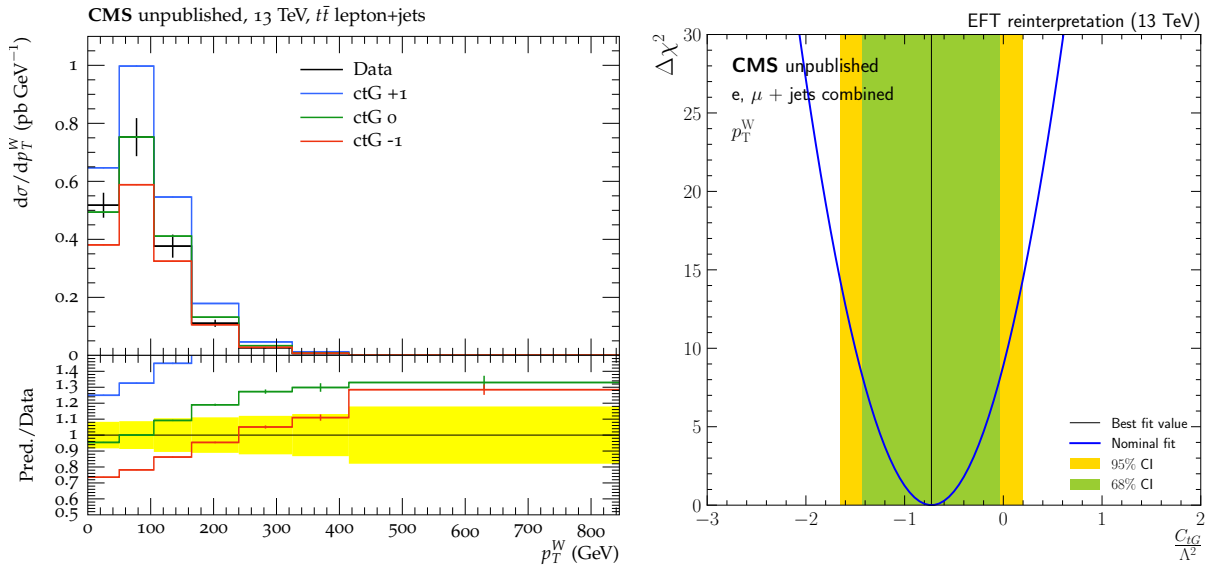


Figure 135: The left panel shows the predictions of EFT simulations using $C_{tG}/\Lambda_{\text{NP}}^2$ strengths of -1, 0 and 1 TeV^{-2} for the p_T^W event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing $C_{tG}/\Lambda_{\text{NP}}^2$ strengths.

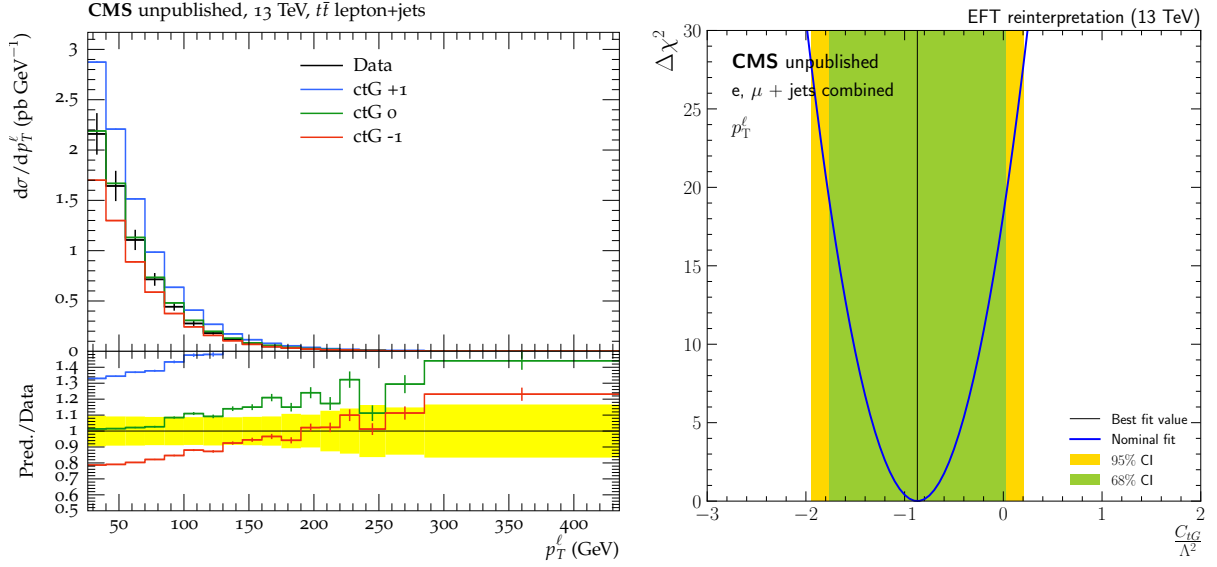


Figure 136: The left panel shows the predictions of EFT simulations using C_{tG}/Λ_{NP}^2 strengths of -1, 0 and 1 TeV $^{-2}$ for the p_T^l event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing C_{tG}/Λ_{NP}^2 strengths.

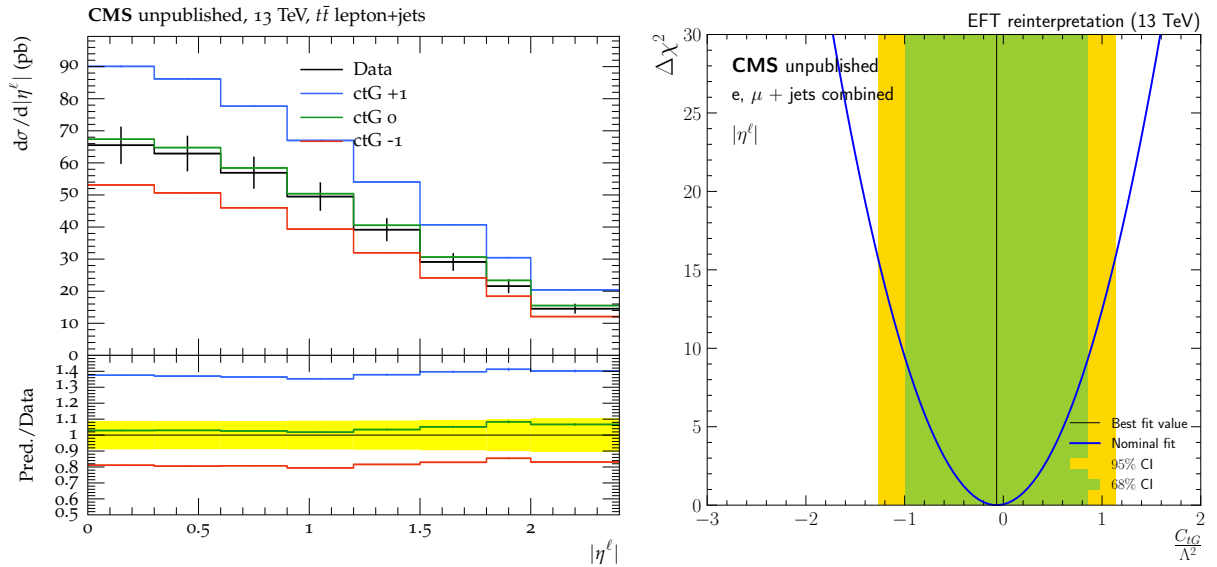


Figure 137: The left panel shows the predictions of EFT simulations using C_{tG}/Λ_{NP}^2 strengths of -1, 0 and 1 TeV $^{-2}$ for the $|\eta^l|$ event variable. The right panel shows nominal fit of the $\Delta\chi^2$ value between the prediction and the absolute $t\bar{t}$ cross section for differing C_{tG}/Λ_{NP}^2 strengths.

Bibliography

- [1] M. Hoch and M. Brice, “CMS experiment”, (2008). CMS Collection.
- [2] T. Mc Cauley, “Collisions recorded by the CMS detector on 7 May 2016 at the start of the year’s physics run”, (2016). CMS Collection.
- [3] CMS Collaboration, “Measurement of differential top quark pair production cross sections in a fiducial volume as a function of event variables in pp collisions at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-15-013, 2015.
- [4] CMS Collaboration, “Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **06** (2018) 002, doi: 10.1007/JHEP06(2018)002, [arXiv:1803.03991](#).
- [5] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi: 10.1103/PhysRevLett.19.1264.
- [6] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi: 10.1016/0029-5582(61)90469-2.
- [7] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc.* **C680519** (1968) 367–377.
- [8] M. Gell-Mann, “A Schematic Model of Baryons and Mesons”, *Phys. Lett.* **8** (1964) 214–215, doi: 10.1016/S0031-9163(64)92001-3.
- [9] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 2”, *Developments in the Quark Theory of Hadrons* **1** (1964) 22–101. Edited by D. Lichtenberg and S. P. Rosen.
- [10] CMS Collaboration, “The CMS experiment at the CERN LHC.”, *JINST* **3** (2008) S08004. 361 p.
- [11] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, “Evidence for the 2π Decay of the K_2^0 Meson”, *Phys. Rev. Lett.* **13** (1964) 138–140, doi: 10.1103/PhysRevLett.13.138.

- [12] Particle Data Group, C. Patrignani et al., “Review of particle physics”, *Chin. Phys. C.* **40** (2016) 100001, doi: 10.1088/1674-1137/40/10/100001.
- [13] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi: 10.1103/PhysRevLett.13.321.
- [14] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi: 10.1103/PhysRevLett.13.508.
- [15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, doi: 10.1103/PhysRevLett.13.585.
- [16] R. P. Feynman, “The Theory of Positrons”, *Phys. Rev.* **76** (1949) 749–759, doi: 10.1103/PhysRev.76.749.
- [17] R. P. Feynman, “Space - time approach to quantum electrodynamics”, *Phys. Rev.* **76** (1949) 769–789, doi: 10.1103/PhysRev.76.769.
- [18] J. Ellis, “TikZ-Feynman: Feynman diagrams with TikZ”, *Comput. Phys. Commun.* **210** (2017) 103–123, doi: 10.1016/j.cpc.2016.08.019, arXiv:1601.05437.
- [19] M. Thomson, “Modern particle physics”. Cambridge University Press, New York, 2013. doi: 10.1017/CBO9781139525367.
- [20] E. Noether, “Invariant Variation Problems”, *Transport Theory Statistics* (1971) 186–207, doi: 10.1080/00411457108231446, arXiv:physics/0503066. Translated by M. Tavel.
- [21] CMS Collaboration, “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV”, *JHEP* **03** (2017) 156, doi: 10.1007/JHEP03(2017)156, arXiv:1609.05331.
- [22] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (1963) 531–533, doi: 10.1103/PhysRevLett.10.531.
- [23] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, doi: 10.1143/PTP.49.652.
- [24] K. Becker, M. Becker, and J. H. Schwarz, “String Theory and M-Theory: A Modern Introduction”. Cambridge University Press, Cambridge, 2006. doi: 10.1017/CBO9780511816086.
- [25] SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89** (2002) 011301, doi: 10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.
- [26] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi: 10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [27] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr., “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 / $R = 4$ kpc/ to UGC 2885 / $R = 122$ kpc/”, *Astrophys. J.* **238** (1980) 471, doi: 10.1086/158003.

- [28] F. Zwicky, “On the Masses of Nebulae and of Clusters of Nebulae”, *Astrophys. J.* **86** (1937) 217, doi: 10.1086/143864.
- [29] Planck Collaboration, “Planck 2013 results. XVI. Cosmological parameters”, *Astron. Astrophys.* **571** (2014) A16, doi: 10.1051/0004-6361/201321591, arXiv:1303.5076.
- [30] Supernova Search Team Collaboration, “Observational evidence from supernovae for an accelerating universe and a cosmological constant”, *Astron. J.* **116** (1998) 1009–1038, doi: 10.1086/300499, arXiv:astro-ph/9805201.
- [31] M. Dine and A. Kusenko, “The Origin of the matter - antimatter asymmetry”, *Rev. Mod. Phys.* **76** (2003) 1, doi: 10.1103/RevModPhys.76.1, arXiv:hep-ph/0303065.
- [32] J. F. Gunion, “A Simplified Summary of Supersymmetry”, *AIP Conf. Proc.* **397** (1997) 41–64, doi: 10.1063/1.52979, arXiv:hep-ph/9704349.
- [33] CDF Collaboration, “Observation of top quark production in $\bar{p}p$ collisions”, *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi: 10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.
- [34] D0 Collaboration, “Observation of the top quark”, *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi: 10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.
- [35] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi: 10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [36] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77** (2017) 663, doi: 10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.
- [37] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641–653. [*Zh. Eksp. Teor. Fiz.*73,1216(1977)].
- [38] V. N. Gribov and L. N. Lipatov, “Deep inelastic $e p$ scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [*Yad. Fiz.*15,781(1972)].
- [39] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys. B* **126** (1977) 298–318, doi: 10.1016/0550-3213(77)90384-4.
- [40] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi: 10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [41] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, “Hadronic top-quark pair production with NNLL threshold resummation”, *Nucl. Phys. B* **855** (2012) 695, doi: 10.1016/j.nuclphysb.2011.10.021, arXiv:1109.1536.
- [42] M. Cacciari et al., “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation”, *Phys. Lett. B* **710** (2012) 612, doi: 10.1016/j.physletb.2012.03.013, arXiv:1111.5869.
- [43] P. Bärnreuther, M. Czakon, and A. Mitov, “Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$ ”, *Phys. Rev. Lett.* **109** (2012) 132001, doi: 10.1103/PhysRevLett.109.132001, arXiv:1204.5201.

- [44] M. Czakon and A. Mitov, “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels”, *JHEP* **12** (2012) 054, doi: 10.1007/JHEP12(2012)054, [arXiv:1207.0236](#).
- [45] M. Czakon and A. Mitov, “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction”, *JHEP* **01** (2013) 080, doi: 10.1007/JHEP01(2013)080, [arXiv:1210.6832](#).
- [46] M. Czakon, P. Fiedler, and A. Mitov, “Total top-quark pair-production cross section at hadron colliders through $O(\alpha_s^4)$ ”, *Phys. Rev. Lett.* **110** (2013) 252004, doi: 10.1103/PhysRevLett.110.252004, [arXiv:1303.6254](#).
- [47] M. Czakon and A. Mitov, “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi: 10.1016/j.cpc.2014.06.021, [arXiv:1112.5675](#).
- [48] “LHCTopWG: Summary Plots”.
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots>.
- [49] L. Asquith et al., “Jet Substructure at the Large Hadron Collider : Experimental Review”, [arXiv:1803.06991](#).
- [50] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section using events in the $e\mu$ final state in pp collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **77** (2017) 172, doi: 10.1140/epjc/s10052-017-4718-8, [arXiv:1611.04040](#).
- [51] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **09** (2017) 051, doi: 10.1007/JHEP09(2017)051, [arXiv:1701.06228](#).
- [52] CMS Collaboration, “First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV”, *Phys. Lett. B* **695** (2011) 424–443, doi: 10.1016/j.physletb.2010.11.058, [arXiv:1010.5994](#).
- [53] CMS Collaboration, “Measurement of the Top-antitop Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV using the Kinematic Properties of Events with Leptons and Jets”, *Eur. Phys. J. C* **71** (2011) 1721, doi: 10.1140/epjc/s10052-011-1721-3, [arXiv:1106.0902](#).
- [54] CMS Collaboration, “Measurement of the $t\bar{t}$ Production Cross Section in pp Collisions at 7 TeV in Lepton + Jets Events Using b -quark Jet Identification”, *Phys. Rev. D* **84** (2011) 092004, doi: 10.1103/PhysRevD.84.092004, [arXiv:1108.3773](#).
- [55] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section and the top quark mass in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **07** (2011) 049, doi: 10.1007/JHEP07(2011)049, [arXiv:1105.5661](#).
- [56] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states”, *Phys. Lett. B* **720** (2013) 83–104, doi: 10.1016/j.physletb.2013.02.021, [arXiv:1212.6682](#).
- [57] CMS Collaboration, “Measurement of the top-antitop production cross section in the tau+jets channel in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **73** (2013) 2386, doi: 10.1140/epjc/s10052-013-2386-x, [arXiv:1301.5755](#).

- [58] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **11** (2012) 067, doi: 10.1007/JHEP11(2012)067, [arXiv:1208.2671](#).
- [59] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **02** (2014) 024, doi: 10.1007/JHEP02(2014)024, 10.1007/JHEP02(2014)102, [arXiv:1312.7582](#).
- [60] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV in dilepton final states containing one τ lepton”, *Phys. Lett. B* **739** (2014) 23–43, doi: 10.1016/j.physletb.2014.10.032, [arXiv:1407.6643](#).
- [61] CMS Collaboration, “Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **76** (2016) 379, doi: 10.1140/epjc/s10052-016-4105-x, [arXiv:1510.03072](#).
- [62] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the $e\mu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **08** (2016) 029, doi: 10.1007/JHEP08(2016)029, [arXiv:1603.02303](#).
- [63] CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **76** (2016) 128, doi: 10.1140/epjc/s10052-016-3956-5, [arXiv:1509.06076](#).
- [64] CMS Collaboration, “Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton”, *JHEP* **03** (2018) 115, doi: 10.1007/JHEP03(2018)115, [arXiv:1711.03143](#).
- [65] CMS Collaboration, “Measurement of the semileptonic $t\bar{t} + \gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **10** (2017) 006, doi: 10.1007/JHEP10(2017)006, [arXiv:1706.08128](#).
- [66] CMS Collaboration, “Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **08** (2018) 011, doi: 10.1007/JHEP08(2018)011, [arXiv:1711.02547](#).
- [67] CMS Collaboration, “Observation of $t\bar{t}H$ production”, *Phys. Rev. Lett.* **120** (2018) 231801, doi: 10.1103/PhysRevLett.120.231801, 10.1130/PhysRevLett.120.231801, [arXiv:1804.02610](#).
- [68] CMS Collaboration, “Observation of top quark production in proton-nucleus collisions”, *Phys. Rev. Lett.* **119** (2017) 242001, doi: 10.1103/PhysRevLett.119.242001, [arXiv:1709.07411](#).
- [69] ATLAS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the τ +jets channel using the ATLAS detector”, *Eur. Phys. J. C* **73** (2013) 2328, doi: 10.1140/epjc/s10052-013-2328-7, [arXiv:1211.7205](#).
- [70] ATLAS Collaboration, “Measurement of the top quark pair cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV using final states with an electron or a muon and a hadronically decaying τ lepton”, *Phys. Lett. B* **717** (2012) 89–108, doi: 10.1016/j.physletb.2012.09.032, [arXiv:1205.2067](#).

- [71] ATLAS Collaboration, “Measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using final states with two high-pt leptons”, *JHEP* **05** (2012) 059, doi: 10.1007/JHEP05(2012)059, [arXiv:1202.4892](#).
- [72] ATLAS Collaboration, “Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel”, *Phys. Lett. B* **711** (2012) 244–263, doi: 10.1016/j.physletb.2012.03.083, [arXiv:1201.1889](#).
- [73] ATLAS Collaboration, “Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states with ATLAS”, *Phys. Lett. B* **707** (2012) 459–477, doi: 10.1016/j.physletb.2011.12.055, [arXiv:1108.3699](#).
- [74] ATLAS Collaboration, “Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **71** (2011) 1577, doi: 10.1140/epjc/s10052-011-1577-6, [arXiv:1012.1792](#).
- [75] ATLAS Collaboration, “Measurement of the inclusive and fiducial $t\bar{t}$ production cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **78** (2018) 487, doi: 10.1140/epjc/s10052-018-5904-z, [arXiv:1712.06857](#).
- [76] ATLAS Collaboration, “Measurement of the $t\bar{t}$ production cross section in the $\tau +$ jets final state in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, *Phys. Rev. D* **95** (2017) 072003, doi: 10.1103/PhysRevD.95.072003, [arXiv:1702.08839](#).
- [77] ATLAS Collaboration, “Measurement of the top pair production cross section in 8 TeV proton-proton collisions using kinematic information in the lepton+jets final state with ATLAS”, *Phys. Rev. D* **91** (2015) 112013, doi: 10.1103/PhysRevD.91.112013, [arXiv:1504.04251](#).
- [78] ATLAS Collaboration, “Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector”, *Eur. Phys. J. C* **74** (2014) 3109, doi: 10.1140/epjc/s10052-016-4501-2, 10.1140/epjc/s10052-014-3109-7, [arXiv:1406.5375](#). [Addendum: *Eur. Phys. J. C* 76,no.11,642(2016)].
- [79] ATLAS Collaboration, “Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector”, *Phys. Lett. B* **761** (2016) 136–157, doi: 10.1016/j.physletb.2016.08.019, 10.1016/j.physletb.2017.09.027, [arXiv:1606.02699](#). [Erratum: *Phys. Lett. B* 772,879(2017)].
- [80] LHCb Collaboration, “First observation of top quark production in the forward region”, *Phys. Rev. Lett.* **115** (2015) 112001, doi: 10.1103/PhysRevLett.115.112001, [arXiv:1506.00903](#).
- [81] CDF, D0 Collaboration, “Combination of measurements of the top-quark pair production cross section from the Tevatron Collider”, *Phys. Rev. D* **89** (2014) 072001, doi: 10.1103/PhysRevD.89.072001, [arXiv:1309.7570](#).
- [82] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

- [83] CMS Collaboration, “Measurement of differential top-quark pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **73** (2013) 2339, doi: 10.1140/epjc/s10052-013-2339-4, [arXiv:1211.2220](#).
- [84] CMS Collaboration, “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **75** (2015) 542, doi: 10.1140/epjc/s10052-015-3709-x, [arXiv:1505.04480](#).
- [85] CMS Collaboration, “Measurement of the integrated and differential $t\bar{t}$ production cross sections for high- p_T top quarks in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. D* **94** (2016) 072002, doi: 10.1103/PhysRevD.94.072002, [arXiv:1605.00116](#).
- [86] CMS Collaboration, “Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions”, *Eur. Phys. J. C* **77** (2017) 459, doi: 10.1140/epjc/s10052-017-4984-5, [arXiv:1703.01630](#).
- [87] CMS Collaboration, “Measurement of normalized differential $t\bar{t}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **04** (2018) 060, doi: 10.1007/JHEP04(2018)060, [arXiv:1708.07638](#).
- [88] CMS Collaboration, “Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV”, *Phys. Rev. D* **95** (2017) 092001, doi: 10.1103/PhysRevD.95.092001, [arXiv:1610.04191](#).
- [89] CMS Collaboration, “Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **776** (2018) 355–378, doi: 10.1016/j.physletb.2017.11.043, [arXiv:1705.10141](#).
- [90] CMS Collaboration, “Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **97** (2018) 112003, doi: 10.1103/PhysRevD.97.112003, [arXiv:1803.08856](#).
- [91] ATLAS Collaboration, “Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector”, *JHEP* **06** (2015) 100, doi: 10.1007/JHEP06(2015)100, [arXiv:1502.05923](#).
- [92] ATLAS Collaboration, “Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”, *Phys. Rev. D* **90** (2014) 072004, doi: 10.1103/PhysRevD.90.072004, [arXiv:1407.0371](#).
- [93] ATLAS Collaboration, “Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **73** (2013) 2261, doi: 10.1140/epjc/s10052-012-2261-1, [arXiv:1207.5644](#).
- [94] ATLAS Collaboration, “Measurement of lepton differential distributions and the top quark mass in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **77** (2017) 804, doi: 10.1140/epjc/s10052-017-5349-9, [arXiv:1709.09407](#).

- [95] ATLAS Collaboration, “Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS”, *Phys. Rev. D* **94** (2016) 092003, doi: 10.1103/PhysRevD.94.092003, arXiv:1607.07281.
- [96] ATLAS Collaboration, “Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, *Eur. Phys. J. C* **76** (2016) 538, doi: 10.1140/epjc/s10052-016-4366-4, arXiv:1511.04716.
- [97] ATLAS Collaboration, “Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS detector”, *Phys. Rev. D* **93** (2016) 032009, doi: 10.1103/PhysRevD.93.032009, arXiv:1510.03818.
- [98] ATLAS Collaboration, “Measurements of differential cross sections of top quark pair production in association with jets in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”, arXiv:1802.06572. Submitted to JHEP.
- [99] ATLAS Collaboration, “Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”, *Phys. Rev. D* **98** (2018) 012003, doi: 10.1103/PhysRevD.98.012003, arXiv:1801.02052.
- [100] ATLAS Collaboration, “Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”, *JHEP* **11** (2017) 191, doi: 10.1007/JHEP11(2017)191, arXiv:1708.00727.
- [101] ATLAS Collaboration, “Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”, *Eur. Phys. J. C* **77** (2017) 292, doi: 10.1140/epjc/s10052-017-4821-x, arXiv:1612.05220.
- [102] ATLAS Collaboration, “Measurement of jet activity produced in top-quark events with an electron, a muon and two b -tagged jets in the final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **77** (2017) 220, doi: 10.1140/epjc/s10052-017-4766-0, arXiv:1610.09978.
- [103] CMS Collaboration, “Measurement of the differential cross sections for top quark pair production as a function of kinematic event variables in pp collisions at $\sqrt{s}=7$ and 8 TeV”, *Phys. Rev. D* **94** (2016) 052006, doi: 10.1103/PhysRevD.94.052006, arXiv:1607.00837.
- [104] F. Marcastel, “CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN”, Technical Report OPEN-PHO-CHART-2013-001, 2013.
- [105] “The Large Hadron Collider”. <https://timeline.web.cern.ch/timeline-header/93>.
- [106] A. Team, “Diagram of an LHC dipole magnet. Schéma d’un aimant dipôle du LHC”, (1999).
- [107] “Public CMS Data Quality Information”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality>.
- [108] “CMS Luminosity - Public Results”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.

- [109] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2018) 161, doi: 10.1007/JHEP07(2018)161, arXiv:1802.02613.
- [110] T. Sakuma and T. McCauley, “Detector and Event Visualization with SketchUp at the CMS Experiment”, *Journal of Physics: Conference Series* **513** (2014) 022032.
- [111] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi: 10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [112] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, Geneva, 2006.
- [113] CMS Collaboration, “Precision crystal calorimetry in LHC Run II with the CMS ECAL”, *JINST* **12** (2017) C01069, doi: 10.1088/1748-0221/12/01/C01069.
- [114] CMS Collaboration, “The CMS-HF quartz fiber calorimeters”, *Journal of Physics: Conference Series* **160** (2009) 012014.
- [115] S. Abdullin et al., “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J. C* **60** (2009) 359–373, doi: 10.1140/epjc/s10052-009-0959-5, 10.1140/epjc/s10052-009-1024-0.
- [116] CMS Collaboration, “The CMS muon system in Run2: preparation, status and first results”, *PoS EPS-HEP2015* (2015) 237, arXiv:1510.05424.
- [117] CMS Collaboration, “The CMS Level-1 trigger system for LHC Run II”, *JINST* **12** (2017) C03021, doi: 10.1088/1748-0221/12/03/C03021.
- [118] CMS Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”, Technical Report CERN-LHCC-2013-011. CMS-TDR-12, 2013.
- [119] “Sherpa and Open Science Grid: Predicting the emergence of jets”. <https://scienode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php>.
- [120] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi: 10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [121] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi: 10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [122] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi: 10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [123] T. Sjöstrand, “The Development of MPI Modelling in PYTHIA”, arXiv:1706.02166.
- [124] Höche, Stefan, “Introduction to parton-shower event generators”, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pp. 235–295. 2015. arXiv:1411.4085. doi: 10.1142/9789814678766_0005.

- [125] J. C. Collins, “Sudakov form-factors”, *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 573–614, doi: 10.1142/9789814503266_0006, [arXiv:hep-ph/0312336](#).
- [126] Sjöstrand, T. and Skands, Peter Z., “Transverse-momentum-ordered showers and interleaved multiple interactions”, *Eur. Phys. J. C* **39** (2005) 129–154, doi: 10.1140/epjc/s2004-02084-y, [arXiv:hep-ph/0408302](#).
- [127] G. Marchesini and B. R. Webber, “Simulation of QCD Jets Including Soft Gluon Interference”, *Nucl. Phys. B* **238** (1984) 1–29, doi: 10.1016/0550-3213(84)90463-2.
- [128] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, doi: 10.1016/j.cpc.2015.01.024, [arXiv:1410.3012](#).
- [129] M. Bahr et al., “Herwig++ physics and manual”, *Eur. Phys. J. C* **58** (2008) 639, doi: 10.1140/epjc/s10052-008-0798-9, [arXiv:0803.0883](#).
- [130] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473–500, doi: 10.1140/epjc/s10052-007-0490-5, [arXiv:0706.2569](#).
- [131] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi: 10.1007/JHEP12(2012)061, [arXiv:1209.6215](#).
- [132] G. Heinrich et al., “NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers”, *JHEP* **08** (2017) 088, doi: 10.1007/JHEP08(2017)088, [arXiv:1703.09252](#).
- [133] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO Higgs boson production via gluon fusion matched with shower in POWHEG”, *JHEP* **04** (2009) 002, doi: 10.1088/1126-6708/2009/04/002, [arXiv:0812.0578](#).
- [134] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145, doi: 10.1016/0370-1573(83)90080-7.
- [135] Sjöstrand, Torbjörn, “Status and developments of event generators”, *PoS LHCP2016* (2016) 007, [arXiv:1608.06425](#).
- [136] Argyropoulos, Spyros and Sjöstrand, Torbjörn, “Effects of color reconnection on $t\bar{t}$ final states at the LHC”, *JHEP* **11** (2014) 043, doi: 10.1007/JHEP11(2014)043, [arXiv:1407.6653](#).
- [137] J. R. Christiansen and P. Z. Skands, “String Formation Beyond Leading Colour”, *JHEP* **08** (2015) 003, doi: 10.1007/JHEP08(2015)003, [arXiv:1505.01681](#).
- [138] “Event generation and parton shower”.
<https://www.gk-eichtheorien.physik.uni-mainz.de/Dateien/Zeppenfeld-3.pdf>.
- [139] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi: 10.1016/S0168-9002(03)01368-8.
- [140] S. Frixione, P. Nason, and G. Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* **09** (2007) 126, doi: 10.1088/1126-6708/2007/09/126, [arXiv:0707.3088](#).
- [141] S. Gieseke, C. Rohr, and A. Siodmok, “Colour reconnections in Herwig++”, *Eur. Phys. J. C* **72** (2012) 2225, doi: 10.1140/epjc/s10052-012-2225-5, [arXiv:1206.0041](#).

- [142] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi: 10.1140/epjc/s10052-016-3988-x, [arXiv:1512.00815](#).
- [143] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi: 10.1007/JHEP04(2015)040, [arXiv:1410.8849](#).
- [144] R. Frederix, E. Re, and P. Torrielli, “Single-top t -channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO”, *JHEP* **09** (2012) 130, doi: 10.1007/JHEP09(2012)130, [arXiv:1207.5391](#).
- [145] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547, doi: 10.1140/epjc/s10052-011-1547-z, [arXiv:1009.2450](#).
- [146] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR”, *Comput. Phys. Commun.* **182** (2011) 1034, doi: 10.1016/j.cpc.2010.12.040, [arXiv:1007.1327](#).
- [147] P. Kant et al., “HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74, doi: 10.1016/j.cpc.2015.02.001, [arXiv:1406.4403](#).
- [148] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev. D* **86** (2012) 094034, doi: 10.1103/PhysRevD.86.094034, [arXiv:1208.5967](#).
- [149] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi: 10.1088/1748-0221/12/10/P10003, [arXiv:1706.04965](#).
- [150] Frühwirth, R., “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444–450, doi: 10.1016/0168-9002(87)90887-4.
- [151] CMS Collaboration, “Adaptive Vertex Fitting”, CMS Note CMS-NOTE-2007-008, 2007.
- [152] CMS Collaboration, “CMS Tracking Performance Results from early LHC Operation”, *Eur. Phys. J. C* **70** (2010) 1165–1192, doi: 10.1140/epjc/s10052-010-1491-3, [arXiv:1007.1988](#).
- [153] CMS Collaboration, “Studies of Tracker Material”, CMS Physics Analysis Summary CMS-PAS-TRK-10-003, 2010.
- [154] CMS Collaboration, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC”, CMS Note CMS-NOTE-2005-001, 2005.
- [155] Y. Coadou, “Boosted Decision Trees and Applications”, *EPJ Web Conf.* **55** (2013) 02004, doi: 10.1051/epjconf/20135502004.
- [156] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm”, *JHEP* **04** (2008) 063, doi: 10.1088/1126-6708/2008/04/063, [arXiv:0802.1189](#).
- [157] CMS Collaboration, “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016”, CMS Detector Performance CMS-DP-2018-028, 2018.

- [158] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P06005, doi: 10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [159] CMS Collaboration, “Muon Identification and Isolation efficiency on full 2016 dataset”, CMS Detector Performance CMS-DP-2017-007, 2017.
- [160] CMS Collaboration, “Jet Performance in pp Collisions at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-JME-10-003, 2010.
- [161] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi: 10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [162] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, doi: 10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [163] CMS Collaboration, “Electron and photon performance in CMS with the full 2016 data sample”, CMS Detector Performance CMS-DP-2017-004, 2017.
- [164] CMS Collaboration, “Muon HLT performance on 2016 data”, CMS Detector Performance CMS-DP-2017-056, 2017.
- [165] CMS Collaboration, “Measurements of differential cross sections for $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons”, CMS Physics Analysis Summary CMS-PAS-TOP-17-014, 2018.
- [166] CMS Collaboration, “Measurement of the differential cross section for $t\bar{t}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-16-011, 2016.
- [167] A. Buckley et al., “Rivet user manual”, *Comput. Phys. Commun.* **184** (2013) 2803–2819, doi: 10.1016/j.cpc.2013.05.021, arXiv:1003.0694.
- [168] CMS Collaboration, “Object definitions for top quark analyses at the particle level”, CMS Note CMS-NOTE-2017-004, 2017.
- [169] G. D’Agostini, “A Multidimensional unfolding method based on Bayes’ theorem”, *Nucl. Instrum. Meth. A* **362** (1995) 487–498, doi: 10.1016/0168-9002(95)00274-X.
- [170] S. Schmitt, “TUnfold: an algorithm for correcting migration effects in high energy physics”, *JINST* **7** (2012) T10003, doi: 10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.
- [171] A. N. Tikhonov, “On the solution of ill-posed problems and the method of regularization”, *Dokl. Akad. Nauk SSSR* **151** (1963) 501–504.
- [172] V. A. Il’in, “Tikhonov’s work on methods of solving ill-posed problems”, *Russ. Math. Surv.* **22** (1967) 142.
- [173] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

- [174] CMS Collaboration, “Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **772** (2017) 752, doi: 10.1016/j.physletb.2017.07.047, [arXiv:1610.00678](#).
- [175] CMS Collaboration, “Measurement of the production cross section of a W boson in association with two b jets in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **77** (2017) 92, doi: 10.1140/epjc/s10052-016-4573-z, [arXiv:1608.07561](#).
- [176] CMS Collaboration, “Measurements of the associated production of a Z boson and b jets in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **77** (2017) 751, doi: 10.1140/epjc/s10052-017-5140-y, [arXiv:1611.06507](#).
- [177] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, doi: 10.1140/epjc/s10052-014-3024-y, [arXiv:1404.5630](#).
- [178] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, “Scaling violations in inclusive e^+e^- annihilation spectra”, *Phys. Rev. D* **27** (1983) 105, doi: 10.1103/PhysRevD.27.105.
- [179] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, “N-Jettiness: An Inclusive Event Shape to Veto Jets”, *Phys. Rev. Lett.* **105** (2010) 092002, doi: 10.1103/PhysRevLett.105.092002, [arXiv:1004.2489](#).
- [180] I. Brivio and M. Trott, “The Standard Model as an Effective Field Theory”, [arXiv:1706.08945](#).
- [181] C. Zhang and S. Willenbrock, “Effective-Field-Theory Approach to Top-Quark Production and Decay”, *Phys. Rev. D* **83** (2011) 034006, doi: 10.1103/PhysRevD.83.034006, [arXiv:1008.3869](#).
- [182] D. Barducci et al., “Interpreting top-quark LHC measurements in the standard-model effective field theory”, [arXiv:1802.07237](#).
- [183] D. Buarque Franzosi and C. Zhang, “Probing the top-quark chromomagnetic dipole moment at next-to-leading order in QCD”, *Phys. Rev. D* **91** (2015) 114010, doi: 10.1103/PhysRevD.91.114010, [arXiv:1503.08841](#).

Acknowledgements

It's finally time to write the acknowledgements and I have spent a lot of time wondering what to say. There have been so many people who have helped me through these last few years, it would be impossible to list them all without writing another huge tome. Nevertheless, I will try.

My first thanks go out to Greg Heath, Freya Blekman and Joel Goldstein, without whom none of this would have been possible. Thank you for the years of guidance, both within the Ph.D. work and outside, you have given me. I would like to thank my doctoral examination jury, Jonas Rademacker, Jorgen D'Hondt, Rebeca Gonzalez Suarez and Miriam Watson. I was unsure what to expect when I walked into my private defence, but it certainly was not a table laden with food and a convivial atmosphere! Thank you for all the interesting discussions we had and the opportunities to make this piece of work even better. I would also like to thank Emyr and Luke for always being there to explain complicated concepts to me and to help me understand what-on-earth is going on with my code! Thanks also to Luke and Winnie for keeping the Bristol computing cluster running smoothly and fixing any issues that appear almost instantaneously. It is amazing to have people like you in a field so heavily reliant on computing.

Next, I want to thank those PartiCOOL physicists with whom I started my Ph.D. Emma, Davide, Leonor, Chiara and Gabby, you made my first year, and subsequent visits, so much fun. Over the years, you have introduced me to the joys of absurdly strong coffee, a whiteboard that is never cleaned and a camaraderie few can surpass. I will never forget the great time we had down in Cornwall together, nor the saga of the wooden spoon!

Moving on to my times spent at CERN. What an incredible place! Thank you to Fionn and Alexander, you were the heart and soul of the office in building 40 (I'll let you figure out which one's which). Kae and Nima, you were always there for me and I won't forget the time we spent together. It is a privilege to know all of you.

The final stage of my Ph.D. lead to Brussels. Wow. What an experience. I still remember walking through those double doors and thinking 'Blimey, this is a big office'. During my stay, this 'big office' has become like a second family to me. Dom, Lieselotte, Isis, Isabelle, Emil, Petra, Seth, Leonidas, Simon, Kirill, Denys, Quentin, Kevin, Shimaa, Giannis, and AR, thank you for making me feel at home, forgiving my truly terrible puns and the many, but not all, ridiculous conversations made over lunch and tea breaks. Thank you for all the evenings spent together, board gaming, eating out, and special events that make life truly worth living.

To Marleen and Martisse, thank you for all of the hard work that you do behind the scenes. I know it means a lot to me and I can safely assume that the rest of the IIHE/Bristol particle

group think that too.

I want to thank all the people on the outside too. There are far too many to list everyone here, but I want to give special thanks to Portia, Rory, Henry, Helena, Steve, Iona, Thomas, Michael and Mike. I have known you guys for quite some time, and you have watched me transform over the years into the person I am now. A large part of that is down to you. Thank you for supporting me during my Ph.D, thank you for allowing me to kip in your abodes when I visit, thank you for keeping me (almost) sane and above all thank you for being really awesome guys!

I can not complete my thanks, without mentioning my family. Thank you so much for bearing with me during this trialling period, especially throughout these last few months. You cannot know how much you have helped me in these last few years.

I hope you all look forward to reading the product of my hard work from the last four years in its entirety! It is an honour to know all of you and I shall leave you with a statement (unabashedly coined from Isis):

Ph.inisheD.

